BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 May 2001 (25.05.2001)

PCT

(10) International Publication Number
WO 01/36597 A1

(51) International Patent Classification[7]: C12N 5/04, 5/10, 15/00, 15/09, 15/63, 15/70, 15/74, 15/82, 15/87, C07H 21/02, 21/04, A01H 1/00, 9/00, 11/00

(21) International Application Number: PCT/US00/31344

(22) International Filing Date:
14 November 2000 (14.11.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/166,228	17 November 1999 (17.11.1999)	US
60/197,899	17 April 2000 (17.04.2000)	US
60/227,439	22 August 2000 (22.08.2000)	US

(71) Applicant *(for all designated States except US)*: **MENDEL BIOTECHNOLOGY, INC.** [US/US]; 21375 Cabot Boulevard, Hayward, CA 94541 (US).

(71) Applicants and
(72) Inventors: **CREELMAN, Robert** [US/US]; 2801 Jennifer Drive, Castro Valley, CA 94546 (US). **YU, Guo-Liang** [CN/US]; 242 Gravatt Drive, Berkeley, CA 94705 (US). **ADAM, Luc** [CA/US]; 25800 Industrial Boulevard, L403, Hayward, CA 94545 (US). **RIECHMANN, Jose Luis** [ES/US]; 115 Moss Avenue #308, Oakland, CA 94611 (US). **HEARD, Jacqueline** [US/US]; 810 Guildford Avenue, San Mateo, CA 94402 (US). **SAMAHA, Raymond** [US/US]; 2224 Albert Lane, Capitola, CA 95010 (US). **PILGRIM, Marsha** [US/US]; 790 Saltillo Place, Fremont, CA 94536 (US). **PINEDA, Omaira** [CO/US]; 19563 Helen Place, Castro Valley, CA 94546 (US). **JIANG, Cai-Zhong** [CN/US]; 34495 Heathrow Terrace, Fremont, CA 94555 (US).

(74) Agent: **GUERRERO, Karen**; Mendel Biotechnology, Inc., 21375 Cabot Boulevard, Hayward, CA 94545 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *With international search report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: PLANT BIOCHEMISTRY-RELATED GENES

(57) Abstract: Recombinant polynucleotides and methods for modifying the phenotype of a plant are provided. In particular, the phenotype that is being modified is a plant's biochemical characteristic.

WO 01/36597 A1

PLANT BIOCHEMISTRY-RELATED GENES

RELATED APPLICATION INFORMATION

The present invention claims the benefit from US Provisional Patent Application Serial Nos. 60/166,228 filed November 17, 1999 and 60/197,899 filed April 17, 2000 and "Plant Trait Modification III" filed August 22, 2000.

FIELD OF THE INVENTION

This invention relates to the field of plant biology. More particularly, the present invention pertains to compositions and methods for phenotypically modifying a plant.

BACKGROUND OF THE INVENTION

Transcription factors can modulate gene expression, either increasing or decreasing (inducing or repressing) the rate of transcription. This modulation results in differential levels of gene expression at various developmental stages, in different tissues and cell types, and in response to different exogenous (e.g., environmental) and endogenous stimuli throughout the life cycle of the organism.

Because transcription factors are key controlling elements of biological pathways, altering the expression levels of one or more transcription factors can change entire biological pathways in an organism. For example, manipulation of the levels of selected transcription factors may result in increased expression of economically useful proteins or metabolic chemicals in plants or to improve other agriculturally relevant characteristics. Conversely, blocked or reduced expression of a transcription factor may reduce biosynthesis of unwanted compounds or remove an undesirable trait. Therefore, manipulating transcription factor levels in a plant offers tremendous potential in agricultural biotechnology for modifying a plant's traits.

The present invention provides novel transcription factors useful for modifying a plant's phenotype in desirable ways, such as modifying a plant's biochemical traits.

SUMMARY OF THE INVENTION

In a first aspect, the invention relates to a recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID Nos. 2N, where N=1-22, or a complementary nucleotide sequence thereof; (b) a nucleotide sequence encoding a polypeptide comprising a conservatively substituted variant of a polypeptide of (a); (c) a nucleotide sequence comprising a sequence selected from those of SEQ ID Nos. 2N-1, where N=1-22, or a complementary nucleotide sequence thereof; (d) a nucleotide sequence comprising silent

substitutions in a nucleotide sequence of (c); (e) a nucleotide sequence which hybridizes under stringent conditions over substantially the entire length of a nucleotide sequence of one or more of: (a), (b), (c), or (d); (f) a nucleotide sequence comprising at least 15 consecutive nucleotides of a sequence of any of (a)-(e); (g) a nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide having a biological activity that modifies a plant's biochemical characteristic; (h) a nucleotide sequence having at least 31% sequence identity to a nucleotide sequence of any of (a)-(g); (i) a nucleotide sequence having at least 60% identity sequence identity to a nucleotide sequence of any of (a)-(g); (j) a nucleotide sequence which encodes a polypeptide having at least 31% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-22; (k) a nucleotide sequence which encodes a polypeptide having at least 60% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-22; and (l) a nucleotide sequence which encodes a conserved domain of a polypeptide having at least 65% sequence identity to a conserved domain of a polypeptide of SEQ ID Nos. 2N, where N=1-22. The recombinant polynucleotide may further comprise a constitutive, inducible, or tissue-active promoter operably linked to the nucleotide sequence. The invention also relates to compositions comprising at least two of the above described polynucleotides.

In a second aspect, the invention is an isolated or recombinant polypeptide comprising a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotide described above.

In another aspect, the invention is a transgenic plant comprising one or more of the above described recombinant polynucleotides. In yet another aspect, the invention is a plant with altered expression levels of a polynucleotide described above or a plant with altered expression or activity levels of an above described polypeptide. Further, the invention is a plant lacking a nucleotide sequence encoding a polypeptide described above. The plant may be a soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, or vegetable brassicas plant.

In a further aspect, the invention relates to a cloning or expression vector comprising the isolated or recombinant polynucleotide described above or cells comprising the cloning or expression vector.

In yet a further aspect, the invention relates to a composition produced by incubating a polynucleotide of the invention with a nuclease, a restriction enzyme, a polymerase; a polymerase and a primer; a cloning vector, or with a cell.

Furthermore, the invention relates to a method for producing a plant having a modified biochemical trait. The method comprises altering the expression of an isolated or recombinant polynucleotide of the invention or altering the expression or activity of a polypeptide of the invention in a plant to produce a modified plant, and selecting the modified plant for a modified biochemical trait.

In another aspect, the invention relates to a method of identifying a factor that is modulated by or interacts with a polypeptide encoded by a polynucleotide of the invention. The method comprises expressing a polypeptide encoded by the polynucleotide in a plant; and identifying at least one factor that is modulated by or interacts with the polypeptide. In one embodiment the method for identifying modulating or interacting factors is by detecting binding by the polypeptide to a promoter sequence, or by detecting interactions between an additional protein and the polypeptide in a yeast two hybrid system, or by detecting expression of a factor by hybridization to a microarray, subtractive hybridization or differential display.

In yet another aspect, the invention is a method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide of interest. The method comprises placing the molecule in contact with a plant comprising the polynucleotide or polypeptide encoded by the polynucleotide of the invention and monitoring one or more of the expression level of the polynucleotide in the plant, the expression level of the polypeptide in the plant, and modulation of an activity of the polypeptide in the plant.

In yet another aspect, the invention relates to an integrated system, computer or computer readable medium comprising one or more character strings corresponding to a polynucleotide of the invention, or to a polypeptide encoded by the polynucleotide. The integrated system, computer or computer readable medium may comprise a link between one or more sequence strings to a modified plant biochemical trait.

In yet another aspect, the invention is a method for identifying a sequence similar or homologous to one or more polynucleotides of the invention, or one or more polypeptides encoded by the polynucleotides. The method comprises providing a sequence database; and, querying the sequence database with one or more target sequences corresponding to the one or more polynucleotides or to the one or more polypeptides to identify one or more sequence members of the database that display sequence similarity or homology to one or more of the one or more target sequences.

The method may further comprise of linking the one or more of the polynucleotides of the invention, or encoded polypeptides, to a modified plant biochemical phenotype.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides a table of exemplary polynucleotide and polypeptide sequences of the invention. The table includes from left to right for each sequence: the SEQ ID No., the internal code reference number (GID), whether the sequence is a polynucleotide or polypeptide sequence, and identification of any conserved domains for the polypeptide sequences.

Figure 2 provides a table of exemplary sequences that are homologous to other sequences provided in the Sequence Listing and that are derived from *Arabidopsis thaliana*. The table includes from left to right: the SEQ ID No., the internal code reference number (GID), identification of the homologous sequence, whether the sequence is a polynucleotide or polypeptide sequence, and identification of any conserved domains for the polypeptide sequences.

Figure 3 provides a table of exemplary sequences that are homologous to the sequences provided in Figures 1 and 2 and that are derived from plants other than *Arabidopsis thaliana*. The table includes from left to right: the SEQ ID No., the internal code reference number (GID), the unique GenBank sequence ID No. (NID), the probability that the comparison was generated by chance (P-value), and the species from which the homologous gene was identified.

DETAILED DESCRIPTION

The present invention relates to polynucleotides and polypeptides, e.g. for modifying phenotypes of plants.

In particular, the polynucleotides or polypeptides are useful for modifying traits associated with a plant's biochemical characteristic when the expression levels of the polynucleotides or expression levels or activity levels of the polypeptides are altered.

The polynucleotides of the invention encode plant transcription factors. The plant transcription factors are derived, e.g., from *Arabidopsis thaliana* and can belong, e.g., to one or more of the following transcription factor families: the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) J. Biol. Chem. 379:633-646); the MYB transcription factor family (Martin and Paz-Ares (1997) Trends Genet. 13:67-73); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) J. Biol. Chem. 378:1079-1101); the WRKY protein family (Ishiguro and Nakamura (1994) Mol. Gen. Genet. 244:563-571); the ankyrin-repeat protein family (Zhang et al. (1992) Plant Cell 4:1575-1588); the

miscellaneous protein (MISC) family (Kim et al. (1997) Plant J. 11:1237-1251); the zinc finger protein (Z) family (Klug and Schwabe (1995) FASEB J. *9:* 597-604); the homeobox (HB) protein family (Duboule (1994) Guidebook to the Homeobox Genes, Oxford University Press); the CAAT-element binding proteins (Forsburg and Guarente (1989) Genes Dev. 3:1166-1178); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) Mol. Gen. Genet. 1996 250:7-16); the NAM protein family; the IAA/AUX proteins (Rouse et al. (1998) Science 279:1371–1373); the HLH/MYC protein family (Littlewood et al. (1994) Prot. Profile 1:639-709); the DNA-binding protein (DBP) family (Tucker et al. (1994) EMBO J. 13:2994-3002); the bZIP family of transcription factors (Foster et al. (1994) FASEB J. 8:192-200); the BPF-1 protein (Box P-binding factor) family (da Costa e Silva et al. (1993) Plant J. 4:125-135); and the golden protein (GLD) family (Hall et al. (1998) Plant Cell 10:925-936).

In addition to methods for modifying a plant phenotype by employing one or more polynucleotides and polypeptides of the invention described herein, the polynucleotides and polypeptides of the invention have a variety of additional uses. These uses include their use in the recombinant production (i.e, expression) of proteins; as regulators of plant gene expression, as diagnostic probes for the presence of complementary or partially complementary nucleic acids (including for detection of natural coding nucleic acids); as substrates for further reactions, e.g., mutation reactions, PCR reactions, or the like, of as substrates for cloning e.g., including digestion or ligation reactions, and for identifying exogenous or endogenous modulators of the transcription factors.

DEFINITIONS

A "polynucleotide" is a nucleic acid sequence comprising a plurality of polymerized nucleotide residues, e.g., at least about 15 consecutive polymerized nucleotide residues, optionally at least about 30 consecutive nucleotides, at least about 50 consecutive nucleotides. In many instances, a polynucleotide comprises a nucleotide sequence encoding a polypeptide (or protein) or a domain or fragment thereof. Additionally, the polynucleotide may comprise a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5' or 3' untranslated regions, a reporter gene, a selectable marker, or the like. The polynucleotide can be single stranded or double stranded DNA or RNA. The polynucleotide optionally comprises modified bases or a modified backbone. The polynucleotide can be, e.g., genomic DNA or RNA, a transcript (such as an mRNA), a cDNA, a PCR product, a cloned DNA, a synthetic DNA or RNA, or the like. The polynucleotide can comprise a sequence in either sense or antisense orientations.

A "recombinant polynucleotide" is a polynucleotide that is not in its native state, e.g., the polynucleotide comprises a nucleotide sequence not found in nature, or the polynucleotide is in a context other than that in which it is naturally found, e.g., separated from nucleotide sequences with which it typically is in proximity in nature, or adjacent (or contiguous with) nucleotide sequences with which it typically is not in proximity. For example, the sequence at issue can be cloned into a vector, or otherwise recombined with one or more additional nucleic acid.

An "isolated polynucleotide" is a polynucleotide whether naturally occurring or recombinant, that is present outside the cell in which it is typically found in nature, whether purified or not. Optionally, an isolated polynucleotide is subject to one or more enrichment or purification procedures, e.g., cell lysis, extraction, centrifugation, precipitation, or the like.

A "recombinant polypeptide" is a polypeptide produced by translation of a recombinant polynucleotide. An "isolated polypeptide," whether a naturally occurring or a recombinant polypeptide, is more enriched in (or out of) a cell than the polypeptide in its natural state in a wild type cell, e.g., more than about 5% enriched, more than about 10% enriched, or more than about 20%, or more than about 50%, or more, enriched, i.e., alternatively denoted: 105%, 110%, 120%, 150% or more, enriched relative to wild type standardized at 100%. Such an enrichment is not the result of a natural response of a wild type plant. Alternatively, or additionally, the isolated polypeptide is separated from other cellular components with which it is typically associated, e.g., by any of the various protein purification methods herein.

The term "transgenic plant" refers to a plant that contains genetic material, not found in a wild type plant of the same species, variety or cultivar. The genetic material may include a transgene, an insertional mutagenesis event (such as by transposon or T-DNA insertional mutagenesis), an activation tagging sequence, a mutated sequence, a homologous recombination event or a sequence modified by chimeraplasty. Typically, the foreign genetic material has been introduced into the plant by human manipulation.

A transgenic plant may contain an expression vector or cassette. The expression cassette typically comprises a polypeptide-encoding sequence operably linked (i.e., under regulatory control of) to appropriate inducible or constitutive regulatory sequences that allow for the expression of polypeptide. The expression cassette can be introduced into a plant by transformation or by breeding after transformation of a parent plant. A plant refers to a whole plant as well as to a plant part, such as seed, fruit, leaf, or root, plant tissue, plant cells or any other plant material, e.g., a plant explant, as well as to progeny thereof, and to *in vitro* systems that mimic biochemical or cellular components or processes in a cell.

The phrase "ectopically expression or altered expression" in reference to a polynucleotide indicates that the pattern of expression in, e.g., a transgenic plant or plant tissue, is different from the expression pattern in a wild type plant or a reference plant of the same species. For example, the polynucleotide or polypeptide is expressed in a cell or tissue type other than a cell or tissue type in which the sequence is expressed in the wild type plant, or by expression at a time other than at the time the sequence is expressed in the wild type plant, or by a response to different inducible agents, such as hormones or environmental signals, or at different expression levels (either higher or lower) compared with those found in a wild type plant. The term also refers to altered expression patterns that are produced by lowering the levels of expression to below the detection level or completely abolishing expression. The resulting expression pattern can be transient or stable, constitutive or inducible. In reference to a polypeptide, the term "ectopic expression or altered expression" further may relate to altered activity levels resulting from the interactions of the polypeptides with exogenous or endogenous modulators or from interactions with factors or as a result of the chemical modification of the polypeptides.

The term "fragment" or "domain," with respect to a polypeptide, refers to a subsequence of the polypeptide. In some cases, the fragment or domain, is a subsequence of the polypeptide which performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA binding domain that binds to a DNA promoter region, an activation domain or a domain for protein-protein interactions. Fragments can vary in size from as few as 6 amino acids to the full length of the intact polypeptide, but are preferably at least about 30 amino acids in length and more preferably at least about 60 amino acids in length. In reference to a nucleotide sequence, "a fragment" refers to any subsequence of a polynucleotide, typically, of at least consecutive about 15 nucleotides, preferably at least about 30 nucleotides, more preferably at least about 50, of any of the sequences provided herein.

The term "trait" refers to a physiological, morphological, biochemical or physical characteristic of a plant or particular plant material or cell. In some instances, this characteristic is visible to the human eye, such as seed or plant size, or can be measured by available biochemical techniques, such as the protein, starch or oil content of seed or leaves or by the observation of the expression level of genes, e.g., by employing Northern analysis, RT-PCR, microarray gene expression assays or reporter gene expression systems, or by agricultural observations such as stress tolerance, yield or pathogen tolerance.

"Trait modification" refers to a detectable difference in a characteristic in a plant ectopically expressing a polynucleotide or polypeptide of the present invention relative to a plant not doing so, such as a wild type plant. In some cases, the trait modification can be evaluated quantitatively. For example, the trait modification can entail at least about a 2% increase or decrease in an observed trait (difference), at least a 5% difference, at least about a 10% difference, at least about a 20% difference, at least about a 30%, at least about a 50%, at least about a 70%, or at least about a 100%, or an even greater difference. It is known that there can be a natural variation in the modified trait. Therefore, the trait modification observed entails a change of the normal distribution of the trait in the plants compared with the distribution observed in wild type plant.

Trait modifications of particular interest include those to seed (such as embryo or endosperm), fruit, root, flower, leaf, stem, shoot, seedling or the like, including: enhanced tolerance to environmental conditions including freezing, chilling, heat, drought, water saturation, radiation and ozone; improved tolerance to microbial, fungal or viral diseases; improved tolerance to pest infestations, including nematodes, mollicutes, parasitic higher plants or the like; decreased herbicide sensitivity; improved tolerance of heavy metals or enhanced ability to take up heavy metals; improved growth under poor photoconditions (e.g., low light and/or short day length), or changes in expression levels of genes of interest. Other phenotype that can be modified relate to the production of plant metabolites, such as variations in the production of taxol, tocopherol, tocotrienol, sterols, phytosterols, vitamins, wax monomers, anti-oxidants, amino acids, lignins, cellulose, tannins, prenyllipids (such as chlorophylls and carotenoids), glucosinolates, and terpenoids, enhanced or compositionally altered protein or oil production (especially in seeds), or modified sugar (insoluble or soluble) and/or starch composition. Physical plant characteristics that can be modified include cell development (such as the number of trichomes), fruit and seed size and number, yields of plant parts such as stems, leaves and roots, the stability of the seeds during storage, characteristics of the seed pod (e.g., susceptibility to shattering), root hair length and quantity, internode distances, or the quality of seed coat. Plant growth characteristics that can be modified include growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, biomass or transpiration characteristics, as well as plant architecture characteristics such as apical dominance, branching patterns, number of organs, organ identity, organ shape or size.

POLYPEPTIDES AND POLYNUCLEOTIDES OF THE INVENTION

The present invention provides, among other things, transcription factors (TFs), and transcription factor homologue polypeptides, and isolated or recombinant polynucleotides encoding the polypeptides. These polypeptides and polynucleotides may be employed to modify a plant's biochemical characteristic.

Exemplary polynucleotides encoding the polypeptides of the invention were identified in the *Arabidopsis thaliana* GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. Polynucleotide sequences meeting such criteria were confirmed as transcription factors.

Additional polynucleotides of the invention were identified by screening *Arabidopsis thaliana* and/or other plant cDNA libraries with probes corresponding to known transcription factors under low stringency hybridization conditions. Additional sequences, including full length coding sequences were subsequently recovered by the rapid amplification of cDNA ends (RACE) procedure, using a commercially available kit according to the manufacturer's instructions. Where necessary, multiple rounds of RACE are performed to isolate 5′ and 3′ ends. The full length cDNA was then recovered by a routine end-to-end polymerase chain reaction (PCR) using primers specific to the isolated 5′ and 3′ ends. Exemplary sequences are provided in the Sequence Listing.

The polynucleotides of the invention were ectopically expressed in overexpressor or knockout plants and changes in the biochemical characteristics of the plants were observed. Therefore, the polynucleotides and polypeptides can be employed to improve the biochemical characteristics of plants.

Making polynucleotides

The polynucleotides of the invention include sequences that encode transcription factors and transcription factor homologue polypeptides and sequences complementary thereto, as well as unique fragments of coding sequence, or sequence complementary thereto. Such polynucleotides can be, e.g., DNA or RNA, e.g., mRNA, cRNA, synthetic RNA, genomic DNA, cDNA synthetic DNA, oligonucleotides, etc. The polynucleotides are either double-stranded or single-stranded, and include either, or both sense (i.e., coding) sequences and antisense (i.e., non-coding, complementary) sequences. The polynucleotides include the coding sequence of a transcription factor, or transcription factor homologue polypeptide, in isolation, in combination with additional coding sequences (e.g., a purification tag, a localization signal, as a fusion-

protein, as a pre-protein, or the like), in combination with non-coding sequences (e.g., introns or inteins, regulatory elements such as promoters, enhancers, terminators, and the like), and/or in a vector or host environment in which the polynucleotide encoding a transcription factor or transcription factor homologue polypeptide is an endogenous or exogenous gene.

A variety of methods exist for producing the polynucleotides of the invention. Procedures for identifying and isolating DNA clones are well known to those of skill in the art, and are described in, e.g., Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA ("Berger"); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2000) ("Ausubel").

Alternatively, polynucleotides of the invention, can be produced by a variety of in vitro amplification methods adapted to the present invention by appropriate selection of specific or degenerate primers. Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qbeta-replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger, Sambrook, and Ausubel, as well as Mullis et al., (1987) PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis). Improved methods for cloning in vitro amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references cited therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. *See*, e.g., Ausubel, Sambrook and Berger, *all supra.*

Alternatively, polynucleotides and oligonucleotides of the invention can be assembled from fragments produced by solid-phase synthesis methods. Typically, fragments of up to approximately 100 bases are individually synthesized and then enzymatically or chemically ligated to produce a desired sequence, e.g., a polynucletotide encoding all or part of a transcription factor. For example, chemical synthesis using the phosphoramidite method is described, e.g., by Beaucage et al. (1981) Tetrahedron Letters 22:1859-69; and Matthes et al. (1984) EMBO J. 3:801-5. According to such methods, oligonucleotides are synthesized, purified,

annealed to their complementary strand, ligated and then optionally cloned into suitable vectors. And if so desired, the polynucleotides and polypeptides of the invention can be custom ordered from any of a number of commercial suppliers.

HOMOLOGOUS SEQUENCES

Sequences homologous, i.e., that share significant sequence identity or similarity, to those provided in the Sequence Listing, derived from *Arabidopsis thaliana* or from other plants of choice are also an aspect of the invention. Homologous sequences can be derived from any plant including monocots and dicots and in particular agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn, potato, cotton, rice, oilseed rape (including canola), sunflower, alfalfa, sugarcane and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, brussel sprouts and kohlrabi). Other crops, fruits and vegetables whose phenotype can be changed include barley, rye, millet, sorghum, currant, avocado, citrus fruits such as oranges, lemons, grapefruit and tangerines, artichoke, cherries, nuts such as the walnut and peanut, endive, leek, roots, such as arrowroot, beet, cassava, turnip, radish, yam, and sweet potato, and beans. The homologous sequences may also be derived from woody species, such pine, poplar and eucalyptus.

Transcription factors that are homologous to the listed sequences will typically share at least about 30% amino acid sequence identity. More closely related transcription factors can share at least about 50%, about 60%, about 65%, about 70%, about 75% or about 80% or about 90% or about 95% or about 98% or more sequence identity with the listed sequences. Factors that are most closely related to the listed sequences share, e.g., at least about 85%, about 90% or about 95% or more % sequence identity to the listed sequences. At the nucleotide level, the sequences will typically share at least about 40% nucleotide sequence identity, preferably at least about 50%, about 60%, about 70% or about 80% sequence identity, and more preferably about 85%, about 90%, about 95% or about 97% or more sequence identity to one or more of the listed sequences. The degeneracy of the genetic code enables major variations in the nucleotide sequence of a polynucleotide while maintaining the amino acid sequence of the encoded protein. Conserved domains within a transcription factor family may exhibit a higher degree of sequence

homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity.

Identifying Nucleic Acids by Hybridization

Polynucleotides homologous to the sequences illustrated in the Sequence Listing can be identified, e.g., by hybridization to each other under stringent or under highly stringent conditions. Single stranded polynucleotides hybridize when they associate based on a variety of well characterized physico-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number), as described in more detail in the references cited above.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is about 5°C to 20°C lower than the thermal melting point (Tm) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire cDNA or selected portions, e.g., to a unique subsequence, of the cDNA under wash conditions of 0.2x SSC to 2.0 x SSC, 0.1% SDS at 50-65° C, for example 0.2 x SSC, 0.1% SDS at 65° C. For identification of less closely related homologues washes can be performed at a lower temperature, e.g., 50° C. In general, stringency is increased by raising the wash temperature and/or decreasing the concentration of SSC.

As another example, stringent conditions can be selected such that an oligonucleotide that is perfectly complementary to the coding oligonucleotide hybridizes to the coding oligonucleotide with at least about a 5-10x higher signal to noise ratio than the ratio for hybridization of the perfectly complementary oligonucleotide to a nucleic acid encoding a transcription factor known as of the filing date of the application. Conditions can be selected such that a higher signal to noise ratio is observed in the particular assay which is used, e.g., about 15x, 25x, 35x, 50x or more. Accordingly, the subject nucleic acid hybridizes to the unique coding oligonucleotide with at least a 2x higher signal to noise ratio as compared to hybridization of the coding oligonucleotide to a nucleic acid encoding known polypeptide. Again, higher signal to noise ratios can be selected, e.g., about 5x, 10x, 25x, 35x, 50x or more. The particular signal will

depend on the label used in the relevant assay, e.g., a fluorescent label, a colorimetric label, a radioactive label, or the like.

Alternatively, transcription factor homologue polypeptides can be obtained by screening an expression library using antibodies specific for one or more transcription factors. With the provision herein of the disclosed transcription factor, and transcription factor homologue nucleic acid sequences, the encoded polypeptide(s) can be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the polypeptide(s) in question. Antibodies can also be raised against synthetic peptides derived from transcription factor, or transcription factor homologue, amino acid sequences. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression library produced from the plant from which it is desired to clone additional transcription factor homologues, using the methods described above. The selected cDNAs can be confirmed by sequencing and enzymatic activity.

SEQUENCE VARIATIONS

It will readily be appreciated by those of skill in the art, that any of a variety of polynucleotide sequences are capable of encoding the transcription factors and transcription factor homologue polypeptides of the invention. Due to the degeneracy of the genetic code, many different polynucleotides can encode identical and/or substantially similar polypeptides in addition to those sequences illustrated in the Sequence Listing.

For example, Table 1 illustrates, e.g., that the codons AGC, AGT, TCA, TCC, TCG, and TCT all encode the same amino acid: serine. Accordingly, at each position in the sequence where there is a codon encoding serine, any of the above trinucleotide sequences can be used without altering the encoded polypeptide.

Table 1

Amino acids			Codon					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	TGC	TGT				
Aspartic acid	Asp	D	GAC	GAT				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	TTC	TTT				
Glycine	Gly	G	GGA	GGC	GGG	GGT		
Histidine	His	H	CAC	CAT				
Isoleucine	Ile	I	ATA	ATC	ATT			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	TTA	TTG	CTA	CTC	CTG	CTT
Methionine	Met	M	ATG					
Asparagine	Asn	N	AAC	AAT				
Proline	Pro	P	CCA	CCC	CCG	CCT		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGT
Serine	Ser	S	AGC	AGT	TCA	TCC	TCG	TCT
Threonine	Thr	T	ACA	ACC	ACG	ACT		
Valine	Val	V	GTA	GTC	GTG	GTT		
Tryptophan	Trp	W	TGG					
Tyrosine	Tyr	Y	TAC	TAT				

Sequence alterations that do not change the amino acid sequence encoded by the polynucleotide are termed "silent" variations. With the exception of the codons ATG and TGG, encoding methionine and tryptophan, respectively, any of the possible codons for the same amino acid can be substituted by a variety of techniques, e.g., site-directed mutagenesis, available in the art. Accordingly, any and all such variations of a sequence selected from the above table are a feature of the invention.

In addition to silent variations, other conservative variations that alter one, or a few amino acids in the encoded polypeptide, can be made without altering the function of the polypeptide, these conservative variants are, likewise, a feature of the invention.

For example, substitutions, deletions and insertions introduced into the sequences provided in the Sequence Listing are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed mutagenesis (Wu (ed.) Meth. Enzymol. (1993) vol. 217, Academic Press) or the other methods noted below. Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof can be

combined to arrive at a sequence. The mutations that are made in the polynucleotide encoding the transcription factor should not place the sequence out of reading frame and should not create complementary regions that could produce secondary mRNA structure. Preferably, the polypeptide encoded by the DNA performs the desired function.

Conservative substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 2 when it is desired to maintain the activity of the protein. Table 2 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as conservative substitutions.

Table 2

Residue	Conservative Substitutions
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Gln	Asn
Cys	Ser
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu, Val
Leu	Ile; Val
Lys	Arg; Gln
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr; Gly
Thr	Ser;Val
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Substitutions that are less conservative than those in Table 2 can be selected by picking residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

FURTHER MODIFYING SEQUENCES OF THE INVENTION—MUTATION/ FORCED EVOLUTION

In addition to generating silent or conservative substitutions as noted, above, the present invention optionally includes methods of modifying the sequences of the Sequence Listing. In the methods, nucleic acid or protein modification methods are used to alter the given sequences to produce new sequences and/or to chemically or enzymatically modify given sequences to change the properties of the nucleic acids or proteins.

Thus, in one embodiment, given nucleic acid sequences are modified, e.g., according to standard mutagenesis or artificial evolution methods to produce modified sequences. For example, Ausubel, *supra*, provides additional details on mutagenesis methods. Artificial forced evolution methods are described, e.g., by Stemmer (1994) Nature 370:389-391, and Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751. Many other mutation and evolution methods are also available and expected to be within the skill of the practitioner.

Similarly, chemical or enzymatic alteration of expressed nucleic acids and polypeptides can be performed by standard methods. For example, sequence can be modified by addition of lipids, sugars, peptides, organic or inorganic compounds, by the inclusion of modified nucleotides or amino acids, or the like. For example, protein modification techniques are illustrated in Ausubel, *supra*. Further details on chemical and enzymatic modifications can be found herein. These modification methods can be used to modify any given sequence, or to modify any sequence produced by the various mutation and artificial evolution modification methods noted herein.

Accordingly, the invention provides for modification of any given nucleic acid by mutation, evolution, chemical or enzymatic modification, or other available methods, as well as

for the products produced by practicing such methods, e.g., using the sequences herein as a starting substrate for the various modification approaches.

For example, optimized coding sequence containing codons preferred by a particular prokaryotic or eukaryotic host can be used e.g., to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced using a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *S. cerevisiae* and mammals are TAA and TGA, respectively. The preferred stop codon for monocotyledonous plants is TGA, whereas insects and *E. coli* prefer to use TAA as the stop codon.

The polynucleotide sequences of the present invention can also be engineered in order to alter a coding sequence for a variety of reasons, including but not limited to, alterations which modify the sequence to facilitate cloning, processing and/or expression of the gene product. For example, alterations are optionally introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to introduce splice sites, etc.

Furthermore, a fragment or domain derived from any of the polypeptides of the invention can be combined with domains derived from other transcription factors or synthetic domains to modify the biological activity of a transcription factor. For instance, a DNA binding domain derived from a transcription factor of the invention can be combined with the activation domain of another transcription factor or with a synthetic activation domain. A transcription activation domain assists in initiating transcription from a DNA binding site. Examples include the transcription activation region of VP16 or GAL4 (Moore et al. (1998) Proc. Natl. Acad. Sci. USA 95: 376-381; and Aoyama et al. (1995) Plant Cell 7:1773-1785), peptides derived from bacterial sequences (Ma and Ptashne (1987) Cell 51; 113-119) and synthetic peptides (Giniger and Ptashne, (1987) Nature 330:670-672).

EXPRESSION AND MODIFICATION OF POLYPEPTIDES

Typically, polynucleotide sequences of the invention are incorporated into recombinant DNA (or RNA) molecules that direct expression of polypeptides of the invention in appropriate host cells, transgenic plants, in vitro translation systems, or the like. Due to the inherent degeneracy of the genetic code, nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can be substituted for any listed sequence to provide for cloning and expressing the relevant homologue.

Vectors, Promoters and Expression Systems

The present invention includes recombinant constructs comprising one or more of the nucleic acid sequences herein. The constructs typically comprise a vector, such as a plasmid, a cosmid, a phage, a virus (e.g., a plant virus), a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), or the like, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

General texts which describe molecular biological techniques useful herein, including the use and production of vectors, promoters and many other relevant topics, include Berger, Sambrook and Ausubel, *supra*. Any of the identified sequences can be incorporated into a cassette or vector, e.g., for expression in plants. A number of expression vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology, Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technology 3: 637-642, for dicotyledonous plants.

Alternatively, non-Ti vectors can be used to transfer the DNA into monocotyledonous plants and cells by using free DNA delivery techniques. Such methods can involve, for example, the use of liposomes, electroporation, microprojectile bombardment, silicon carbide whiskers, and viruses. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9: 957-962) and corn (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technology 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48, and for Agrobacterium-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750).

Typically, plant transformation vectors include one or more cloned plant coding sequence (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start

site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which can be useful for expressing the TF sequence include: the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in most plant tissues (*see,* e.g., Odel et al. (1985) Nature 313:810); the nopaline synthase promoter (An et al. (1988) Plant Physiol 88:547); and the octopine synthase promoter (Fromm et al. (1989) Plant Cell 1: 977).

A variety of plant gene promoters that regulate gene expression in response to environmental, hormonal, chemical, developmental signals, and in a tissue-active manner can be used for expression of a TF sequence in plants. Choice of a promoter is based largely on the phenotype of interest and is determined by such factors as tissue (e.g., seed, fruit, root, pollen, vascular tissue, flower, carpel, etc.), inducibility (e.g., in response to wounding, heat, cold, drought, light, pathogens, etc.), timing, developmental stage, and the like. Numerous known promoters have been characterized and can favorable be employed to promote expression of a polynucleotide of the invention in a transgenic plant or cell of interest. For example, tissue specific promoters include: seed-specific promoters (such as the napin, phaseolin or DC3 promoter described in US Pat. No. 5,773,697), fruit-specific promoters that are active during fruit ripening (such as the dru 1 promoter (US Pat. No. 5,783,393), or the 2A11 promoter (US Pat. No. 4,943,674) and the tomato polygalacturonase promoter (Bird et al. (1988) Plant Mol Biol 11:651), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186, pollen-active promoters such as PTA29, PTA26 and PTA13 (US Pat. No. 5,792,929), promoters active in vascular tissue (Ringli and Keller (1998) Plant Mol Biol 37:977-988), flower-specific (Kaiser et al, (1995) Plant Mol Biol 28:231-243), pollen (Baerson et al. (1994) Plant Mol Biol 26:1947-1959), carpels (Ohl et al. (1990) Plant Cell 2:837-848), pollen and ovules (Baerson et al. (1993) Plant Mol Biol 22:255-267), auxin-inducible promoters (such as that described in van der Kop et al. (1999) Plant Mol Biol 39:979-990 or Baumann et al. (1999) Plant Cell 11:323-334), cytokinin-inducible promoter (Guevara-Garcia (1998) Plant Mol Biol 38:743-753), promoters responsive to gibberellin (Shi et al. (1998) Plant Mol Biol 38:1053-1060, Willmott et al. (1998) 38:817-825) and the like. Additional promoters are those that elicit expression in response to heat (Ainley et al. (1993) Plant Mol Biol 22: 13-23), light (e.g., the pea rbcS-3A promoter, Kuhlemeier et al. (1989) Plant Cell 1:471, and the maize rbcS promoter, Schaffner and Sheen (1991) Plant Cell 3: 997); wounding (e.g., *wunI*, Siebertz et al. (1989) Plant Cell 1: 961); pathogens (such as the PR-1 promoter described in Buchel et al. (1999) Plant Mol. Biol. 40:387-396, and the PDF1.2 promoter described in Manners et al. (1998) Plant Mol. Biol. 38:1071-80),

and chemicals such as methyl jasmonate or salicylic acid (Gatz et al. (1997) Plant Mol Biol 48: 89-108). In addition, the timing of the expression can be controlled by using promoters such as those acting at senescence (An and Amazon (1995) Science 270: 1986-1988); or late seed development (Odell et al. (1994) Plant Physiol 106:447-458).

Plant expression vectors can also include RNA processing signals that can be positioned within, upstream or downstream of the coding sequence. In addition, the expression vectors can include additional regulatory sequences from the 3'-untranslated region of plant genes, e.g., a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Additional Expression Elements

Specific initiation signals can aid in efficient translation of coding sequences. These signals can include, e.g., the ATG initiation codon and adjacent sequences. In cases where a coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence (e.g., a mature protein coding sequence), or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon can be separately provided. The initiation codon is provided in the correct reading frame to facilitate transcription. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of enhancers appropriate to the cell system in use.

Expression Hosts

The present invention also relates to host cells which are transduced with vectors of the invention, and the production of polypeptides of the invention (including fragments thereof) by recombinant techniques. Host cells are genetically engineered (i.e, nucleic acids are introduced, e.g., transduced, transformed or transfected) with the vectors of this invention, which may be, for example, a cloning vector or an expression vector comprising the relevant nucleic acids herein. The vector is optionally a plasmid, a viral particle, a phage, a naked nucleic acids, *etc.* The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the relevant gene. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, Sambrook and Ausubel.

The host cell can be a eukaryotic cell, such as a yeast cell, or a plant cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Plant protoplasts are also suitable for some

applications. For example, the DNA fragments are introduced into plant tissues, cultured plant cells or plant protoplasts by standard methods including electroporation (Fromm et al., (1985) Proc. Natl. Acad. Sci. USA 82, 5824, infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et al., (1982) Molecular Biology of Plant Tumors, (Academic Press, New York) pp. 549-560; US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., (1987) Nature 327, 70-73), use of pollen as vector (WO 85/01856), or use of *Agrobacterium tumefaciens* or *A. rhizogenes* carrying a T-DNA plasmid in which DNA fragments are cloned. The T-DNA plasmid is transmitted to plant cells upon infection by *Agrobacterium tumefaciens*, and a portion is stably integrated into the plant genome (Horsch et al. (1984) Science 233:496-498; Fraley et al. (1983) Proc. Natl. Acad. Sci. USA 80, 4803).

The cell can include a nucleic acid of the invention which encodes a polypeptide, wherein the cells expresses a polypeptide of the invention. The cell can also include vector sequences, or the like.. Furthermore, cells and transgenic plants which include any polypeptide or nucleic acid above or throughout this specification, e.g., produced by transduction of a vector of the invention, are an additional feature of the invention.

For long-term, high-yield production of recombinant proteins, stable expression can be used. Host cells transformed with a nucleotide sequence encoding a polypeptide of the invention are optionally cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein or fragment thereof produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly, depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding mature proteins of the invention can be designed with signal sequences which direct secretion of the mature polypeptides through a prokaryotic or eukaryotic cell membrane.

Modified Amino Acids

Polypeptides of the invention may contain one or more modified amino acids. The presence of modified amino acids may be advantageous in, for example, increasing polypeptide half-life, reducing polypeptide antigenicity or toxicity, increasing polypeptide storage stability, or the like. Amino acid(s) are modified, for example, co-translationally or post-translationally during recombinant production or modified by synthetic or chemical means.

Non-limiting examples of a modified amino acid include incorporation or other use of acetylated amino acids, glycosylated amino acids, sulfated amino acids, prenylated (e.g., farnesylated, geranylgeranylated) amino acids, PEG modified (e.g., "PEGylated") amino acids,

biotinylated amino acids, carboxylated amino acids, phosphorylated amino acids, etc. References adequate to guide one of skill in the modification of amino acids are replete throughout the literature.

IDENTIFICATION OF ADDITIONAL FACTORS

A transcription factor provided by the present invention can also be used to identify additional endogenous or exogenous molecules that can affect a phentoype or trait of interest. On the one hand, such molecules include organic (small or large molecules) and/or inorganic compounds that affect expression of (i.e., regulate) a particular transcription factor. Alternatively, such molecules include endogenous molecules that are acted upon either at a transcriptional level by a transcription factor of the invention to modify a phenotype as desired. For example, the transcription factors can be employed to identify one or more downstream gene with which is subject to a regulatory effect of the transcription factor. In one approach, a transcription factor or transcription factor homologue of the invention is expressed in a host cell, e.g, a transgenic plant cell, tissue or explant, and expression products, either RNA or protein, of likely or random targets are monitored, e.g., by hybridization to a microarray of nucleic acid probes corresponding to genes expressed in a tissue or cell type of interest, by two-dimensional gel electrophoresis of protein products, or by any other method known in the art for assessing expression of gene products at the level of RNA or protein. Alternatively, a transcription factor of the invention can be used to identify promoter sequences (i.e., binding sites) involved in the regulation of a downstream target. After identifying a promoter sequence, interactions between the transcription factor and the promoter sequence can be modified by changing specific nucleotides in the promoter sequence or specific amino acids in the transcription factor that interact with the promoter sequence to alter a plant trait. Typically, transcription factor DNA binding sites are identified by gel shift assays. After identifying the promoter regions, the promoter region sequences can be employed in double-stranded DNA arrays to identify molecules that affect the interactions of the transcription factors with their promoters (Bulyk et al. (1999) Nature Biotechnology 17:573-577).

The identified transcription factors are also useful to identify proteins that modify the activity of the transcription factor. Such modification can occur by covalent modification, such as by phosphorylation, or by protein-protein (homo or-heteropolymer) interactions. Any method suitable for detecting protein-protein interactions can be employed. Among the methods that can be employed are co-immunoprecipitation, cross-linking and co-purification through gradients or chromatographic columns, and the two-hybrid yeast system.

The two-hybrid system detects protein interactions in vivo and is described in Chien, et al., (1991), Proc. Natl. Acad. Sci. USA 88, 9578-9582 and is commercially available from Clontech (Palo Alto, Calif.). In such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain of a transcription activator protein fused to the TF polypeptide and the other consists of the transcription activator protein's activation domain fused to an unknown protein that is encoded by a cDNA that has been recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (e.g., lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product. Then, the library plasmids responsible for reporter gene expression are isolated and sequenced to identify the proteins encoded by the library plasmids. After identifying proteins that interact with the transcription factors, assays for compounds that interfere with the TF protein-protein interactions can be preformed.

IDENTIFICATION OF MODULATORS

In addition to the intracellular molecules described above, extracellular molecules that alter activity or expression of a transcription factor, either directly or indirectly, can be identified. For example, the methods can entail first placing a candidate molecule in contact with a plant or plant cell. The molecule can be introduced by topical administration, such as spraying or soaking of a plant, and then the molecule's effect on the expression or activity of the TF polypeptide or the expression of the polynucleotide monitored. Changes in the expression of the TF polypeptide can be monitored by use of polyclonal or monoclonal antibodies, gel electrophoresis or the like. Changes in the expression of the corresponding polynucleotide sequence can be detected by use of microarrays, Northerns, quantitative PCR, or any other technique for monitoring changes in mRNA expression. These techniques are exemplified in Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1998). Such changes in the expression levels can be correlated with modified plant traits and thus identified molecules can be useful for soaking or spraying on fruit, vegetable and grain crops to modify traits in plants.

Essentially any available composition can be tested for modulatory activity of expression or activity of any nucleic acid or polypeptide herein. Thus, available libraries of compounds such as chemicals, polypeptides, nucleic acids and the like can be tested for modulatory activity.

Often, potential modulator compounds can be dissolved in aqueous or organic (e.g., DMSO-based) solutions for easy delivery to the cell or plant of interest in which the activity of the modulator is to be tested. Optionally, the assays are designed to screen large modulator composition libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays).

In one embodiment, high throughput screening methods involve providing a combinatorial library containing a large number of potential compounds (potential modulator compounds). Such "combinatorial chemical libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as target compounds.

A combinatorial chemical library can be, e.g., a collection of diverse chemical compounds generated by chemical synthesis or biological synthesis. For example, a combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (e.g., in one example, amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound of a set length). Exemplary libraries include peptide libraries, nucleic acid libraries, antibody libraries (see, e.g., Vaughn et al. (1996) Nature Biotechnology, 14(3):309-314 and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al. Science (1996) 274:1520-1522 and U.S. Patent 5,593,853), peptide nucleic acid libraries (see, e.g., U.S. Patent 5,539,083), and small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337) and the like.

Preparation and screening of combinatorial or other libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Patent 5,010,175, Furka, Int. J. Pept. Prot. Res. 37:487-493 (1991) and Houghton et al. Nature 354:84-88 (1991)). Other chemistries for generating chemical diversity libraries can also be used.

In addition, as noted, compound screening equipment for high-throughput screening is generally available, e.g., using any of a number of well known robotic systems that have also been developed for solution phase chemistries useful in assay systems. These systems include automated workstations including an automated synthesis apparatus and robotic systems utilizing robotic arms. Any of the above devices are suitable for use with the present invention, e.g., for

high-throughput screening of potential modulators. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein will be apparent to persons skilled in the relevant art.

Indeed, entire high throughput screening systems are commercially available. These systems typically automate entire procedures including all sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. Similarly, microfluidic implementations of screening are also commercially available.

The manufacturers of such systems provide detailed protocols the various high throughput. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like. The integrated systems herein, in addition to providing for sequence alignment and, optionally, synthesis of relevant nucleic acids, can include such screening apparatus to identify modulators that have an effect on one or more polynucleotides or polypeptides according to the present invention.

In some assays it is desirable to have positive controls to ensure that the components of the assays are working properly. At least two types of positive controls are appropriate. That is, known transcriptional activators or inhibitors can be incubated with cells/plants/ etc. in one sample of the assay, and the resulting increase/decrease in transcription can be detected by measuring the resulting increase in RNA/ protein expression, etc., according to the methods herein. It will be appreciated that modulators can also be combined with transcriptional activators or inhibitors to find modulators which inhibit transcriptional activation or transcriptional repression. Either expression of the nucleic acids and proteins herein or any additional nucleic acids or proteins activated by the nucleic acids or proteins herein, or both, can be monitored.

In an embodiment, the invention provides a method for identifying compositions that modulate the activity or expression of a polynucleotide or polypeptide of the invention. For example, a test compound, whether a small or large molecule, is placed in contact with a cell, plant (or plant tissue or explant), or composition comprising the polynucleotide or polypeptide of interest and a resulting effect on the cell, plant, (or tissue or explant) or composition is evaluated by monitoring, either directly or indirectly, one or more of: expression level of the polynucleotide or polypeptide, activity (or modulation of the activity) of the polynucleotide or polypeptide. In some cases, an alteration in a plant phenotype can be detected following contact of a plant (or

plant cell, or tissue or explant) with the putative modulator, e.g., by modulation of expression or activity of a polynucleotide or polypeptide of the invention.

SUBSEQUENCES

Also contemplated are uses of polynucleotides, also referred to herein as oligonucleotides, typically having at least 12 bases, preferably at least 15, more preferably at least 20, 30, or 50 bases, which hybridize under at least highly stringent (or ultra-high stringent or ultra-ultra- high stringent conditions) conditions to a polynucleotide sequence described above. The polynucleotides may be used as probes, primers, sense and antisense agents, and the like, according to methods as noted *supra*.

Subsequences of the polynucleotides of the invention, including polynucleotide fragments and oligonucleotides are useful as nucleic acid probes and primers. An oligonucleotide suitable for use as a probe or primer is at least about 15 nucleotides in length, more often at least about 18 nucleotides, often at least about 21 nucleotides, frequently at least about 30 nucleotides, or about 40 nucleotides, or more in length. A nucleic acid probe is useful in hybridization protocols, e.g., to identify additional polypeptide homologues of the invention, including protocols for microarray experiments. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods. See Sambrook and Ausubel, *supra*.

In addition, the invention includes an isolated or recombinant polypeptide including a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotides of the invention. For example, such polypeptides, or domains or fragments thereof, can be used as immunogens, e.g., to produce antibodies specific for the polypeptide sequence, or as probes for detecting a sequence of interest. A subsequence can range in size from about 15 amino acids in length up to and including the full length of the polypeptide.

PRODUCTION OF TRANSGENIC PLANTS

Modification of Traits

The polynucleotides of the invention are favorably employed to produce transgenic plants with various traits, or characteristics, that have been modified in a desirable manner, e.g., to improve the seed characteristics of a plant. For example, alteration of expression levels or patterns (e.g., spatial or temporal expression patterns) of one or more of the transcription factors

(or transcription factor homologues) of the invention, as compared with the levels of the same protein found in a wild type plant, can be used to modify a plant's traits. An illustrative example of trait modification, improved biochemical characteristics, by altering expression levels of a particular transcription factor is described further in the Examples and the Sequence Listing.

Antisense and Cosuppression Approaches

In addition to expression of the nucleic acids of the invention as gene replacement or plant phenotype modification nucleic acids, the nucleic acids are also useful for sense and anti-sense suppression of expression, e.g., to down-regulate expression of a nucleic acid of the invention, e.g., as a further mechanism for modulating plant phenotype. That is, the nucleic acids of the invention, or subsequences or anti-sense sequences thereof, can be used to block expression of naturally occurring homologous nucleic acids. A variety of sense and anti-sense technologies are known in the art, e.g., as set forth in Lichtenstein and Nellen (1997) Antisense Technology: A Practical Approach IRL Press at Oxford University, Oxford, England. In general, sense or anti-sense sequences are introduced into a cell, where they are optionally amplified, e.g., by transcription. Such sequences include both simple oligonucleotide sequences and catalytic sequences such as ribozymes.

For example, a reduction or elimination of expression (i.e., a "knock-out") of a transcription factor or transcription factor homologue polypeptide in a transgenic plant, e.g., to modify a plant trait, can be obtained by introducing an antisense construct corresponding to the polypeptide of interest as a cDNA. For antisense suppression, the transcription factor or homologue cDNA is arranged in reverse orientation (with respect to the coding sequence) relative to the promoter sequence in the expression vector. The introduced sequence need not be the full length cDNA or gene, and need not be identical to the cDNA or gene found in the plant type to be transformed. Typically, the antisense sequence need only be capable of hybridizing to the target gene or RNA of interest. Thus, where the introduced sequence is of shorter length, a higher degree of homology to the endogenous transcription factor sequence will be needed for effective antisense suppression. While antisense sequences of various lengths can be utilized, preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides. Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous transcription factor gene in the plant cell.

Suppression of endogenous transcription factor gene expression can also be achieved using a ribozyme. Ribozymes are RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 and U.S. Patent No. 5,543,508. Synthetic ribozyme sequences including antisense RNAs can be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that hybridize to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Vectors in which RNA encoded by a transcription factor or transcription factor homologue cDNA is over-expressed can also be used to obtain co-suppression of a corresponding endogenous gene, e.g., in the manner described in U.S. Patent No. 5,231,020 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire transcription factor cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous transcription factor gene of interest. However, as with antisense suppression, the suppressive efficiency will be enhanced as specificity of hybridization is increased, e.g., as the introduced sequence is lengthened, and/or as the sequence similarity between the introduced sequence and the endogenous transcription factor gene is increased.

Vectors expressing an untranslatable form of the transcription factor mRNA, e.g., sequences comprising one or more stop codon, or nonsense mutation) can also be used to suppress expression of an endogenous transcription factor, thereby reducing or eliminating it's activity and modifying one or more traits. Methods for producing such constructs are described in U.S. Patent No. 5,583,021. Preferably, such constructs are made by introducing a premature stop codon into the transcription factor gene. Alternatively, a plant trait can be modified by gene silencing using double-strand RNA (Sharp (1999) Genes and Development 13: 139-141).

Another method for abolishing the expression of a gene is by insertion mutagenesis using the T-DNA of *Agrobacterium tumefaciens*. After generating the insertion mutants, the mutants can be screened to identify those containing the insertion in a transcription factor or transcription factor homologue gene. Plants containing a single transgene insertion event at the desired gene can be crossed to generate homozygous plants for the mutation (Koncz et al. (1992) Methods in Arabidopsis Research, World Scientific).

Alternatively, a plant phenotype can be altered by eliminating an endogenous gene, such as a transcription factor or transcription factor homologue, e.g., by homologous recombination (Kempin et al. (1997) Nature 389:802).

A plant trait can also be modified by using the cre-lox system (for example, as described in US Pat. No. 5,658,772). A plant genome can be modified to include first and second lox sites

that are then contacted with a Cre recombinase. If the lox sites are in the same orientation, the intervening DNA sequence between the two sites is excised. If the lox sites are in the opposite orientation, the intervening sequence is inverted.

The polynucleotides and polypeptides of this invention can also be expressed in a plant in the absence of an expression cassette by manipulating the activity or expression level of the endogenous gene by other means. For example, by ectopically expressing a gene by T-DNA activation tagging (Ichikawa et al. (1997) Nature 390 698-701; Kakimoto et al. (1996) Science 274: 982-985). This method entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has inserted into the genome, expression of a flanking gene coding sequence becomes deregulated. In another example, the transcriptional machinery in a plant can be modified so as to increase transcription levels of a polynucleotide of the invention (*See*, e.g., PCT Publications WO 96/06166 and WO 98/53057 which describe the modification of the DNA binding specificity of zinc finger proteins by changing particular amino acids in the DNA binding motif).

The transgenic plant can also include the machinery necessary for expressing or altering the activity of a polypeptide encoded by an endogenous gene, for example by altering the phosphorylation state of the polypeptide to maintain it in an activated state.

Transgenic plants (or plant cells, or plant explants, or plant tissues) incorporating the polynucleotides of the invention and/or expressing the polypeptides of the invention can be produced by a variety of well established techniques as described above. Following construction of a vector, most typically an expression cassette, including a polynucleotide, e.g., encoding a transcription factor or transcription factor homologue, of the invention, standard techniques can be used to introduce the polynucleotide into a plant, a plant cell, a plant explant or a plant tissue of interest. Optionally, the plant cell, explant or tissue can be regenerated to produce a transgenic plant.

The plant can be any higher plant, including gymnosperms, monocotyledonous and dicotyledenous plants. Suitable protocols are available for *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip), *Cruciferae* (cabbage, radish, rapeseed, broccoli, etc.), *Curcurbitaceae* (melons and cucumber), *Gramineae* (wheat, corn, rice, barley, millet, etc.), *Solanaceae* (potato, tomato, tobacco, peppers, etc.), and various other crops. See protocols described in Ammirato et al. (1984) Handbook of Plant Cell Culture –Crop Species. Macmillan Publ. Co. Shimamoto et al. (1989) Nature 338:274-276; Fromm et al. (1990) Bio/Technology 8:833-839; and Vasil et al. (1990) Bio/Technology 8:429-434.

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells is now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods can include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumeficiens* mediated transformation. Transformation means introducing a nucleotide sequence in a plant in a manner to cause stable or transient expression of the sequence.

Successful examples of the modification of plant characteristics by transformation with cloned sequences which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include: U.S. Patent Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

Following transformation, plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic or herbicide resistance on the transformed plants, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic or herbicide.

After transformed plants are selected and grown to maturity, those plants showing a modified trait are identified. The modified trait can be any of those traits described above. Additionally, to confirm that the modified trait is due to changes in expression levels or activity of the polypeptide or polynucleotide of the invention can be determined by analyzing mRNA expression using Northern blots, RT-PCR or microarrays, or protein expression using immunoblots or Western blots or gel shift assays.

INTEGRATED SYSTEMS—SEQUENCE IDENTITY

Additionally, the present invention may be an integrated system, computer or computer readable medium that comprises an instruction set for determining the identity of one or more sequences in a database. In addition, the instruction set can be used to generate or identify sequences that meet any specified criteria. Furthermore, the instruction set may be used to associate or link certain functional benefits, such improved biochemical characteristics, with one or more identified sequence.

For example, the instruction set can include, e.g., a sequence comparison or other alignment program, e.g., an available program such as, for example, the Wisconsin Package Version 10.0, such as BLAST, FASTA, PILEUP, FINDPATTERNS or the like (GCG, Madision, WI). Public sequence databases such as GenBank, EMBL, Swiss-Prot and PIR or private sequence databases such as PhytoSeq (Incyte Pharmaceuticals, Palo Alto, CA) can be searched.

Alignment of sequences for comparison can be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. U.S.A. 85: 2444, by computerized implementations of these algorithms. After alignment, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window can be a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 contiguous positions. A description of the method is provided in Ausubel et al., *supra*.

A variety of methods of determining sequence relationships can be used, including manual alignment and computer assisted sequence alignment and analysis. This later approach is a preferred approach in the present invention, due to the increased throughput afforded by computer assisted methods. As noted above, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

One example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al. J. Mol. Biol 215:403-410 (1990). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each

direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see*, e.g., Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence (and, therefore, in this context, homologous) if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, or less than about 0.01, and or even less than about 0.001. An additional example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. The program can align, e.g., up to 300 sequences of a maximum length of 5,000 letters.

The integrated system, or computer typically includes a user input interface allowing a user to selectively view one or more sequence records corresponding to the one or more character strings, as well as an instruction set which aligns the one or more character strings with each other or with an additional character string to identify one or more region of sequence similarity. The system may include a link of one or more character strings with a particular phenotype or gene function. Typically, the system includes a user readable output element which displays an alignment produced by the alignment instruction set.

The methods of this invention can be implemented in a localized or distributed computing environment. In a distributed environment, the methods may implemented on a single computer comprising multiple processors or on a multiplicity of computers. The computers can be linked, e.g. through a common bus, but more preferably the computer(s) are nodes on a network. The network can be a generalized or a dedicated local or wide-area network and, in certain preferred embodiments, the computers may be components of an intra-net or an internet.

Thus, the invention provides methods for identifying a sequence similar or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

Any sequence herein can be entered into the database, before or after querying the database. This provides for both expansion of the database and, if done before the querying step, for insertion of control sequences into the database. The control sequences can be detected by the query to ensure the general integrity of both the database and the query. As noted, the query can be performed using a web browser based interface. For example, the database can be a centralized public database such as those noted herein, and the querying can be done from a remote terminal or computer across an internet or intranet.

EXAMPLES

The following examples are intended to illustrate but not limit the present invention.

EXAMPLE I. FULL LENGTH GENE IDENTIFICATION AND CLONING

Putative transcription factor sequences (genomic or ESTs) related to known transcription factors were identified in the *Arabidopsis thaliana* GenBank database using the tblastn sequence analysis program using default parameters and a P-value cutoff threshold of –4 or –5 or lower, depending on the length of the query sequence. Putative transcription factor sequence hits were then screened to identify those containing particular sequence strings. If the sequence hits contained such sequence strings, the sequences were confirmed as transcription factors.

Alternatively, Arabidopsis *thaliana* cDNA libraries derived from different tissues or treatments, or genomic libraries were screened to identify novel members of a transcription family using a low stringency hybridization approach. Probes were synthesized using gene specific primers in a standard PCR reaction (annealing temperature 60° C) and labeled with ^{32}P dCTP using the High Prime DNA Labeling Kit (Boehringer Mannheim). Purified radiolabelled probes were added to filters immersed in Church hybridization medium (0.5 M $NaPO_4$ pH 7.0, 7% SDS, 1 % w/v bovine serum albumin) and hybridized overnight at 60 °C with shaking. Filters were washed two times for 45 to 60 minutes with 1xSCC, 1% SDS at 60° C.

To identify additional sequence 5' or 3' of a partial cDNA sequence in a cDNA library, 5' and 3' rapid amplification of cDNA ends (RACE) was performed using the Marathon™ cDNA amplification kit (Clontech, Palo Alto, CA). Generally, the method entailed first isolating

poly(A) mRNA, performing first and second strand cDNA synthesis to generate double stranded cDNA, blunting cDNA ends, followed by ligation of the Marathon™ Adaptor to the cDNA to form a library of adaptor-ligated ds cDNA.

Gene-specific primers were designed to be used along with adaptor specific primers for both 5' and 3' RACE reactions. Nested primers, rather than single primers, were used to increase PCR specificity. Using 5' and 3' RACE reactions, 5' and 3' RACE fragments were obtained, sequenced and cloned. The process can be repeated until 5' and 3' ends of the full-length gene were identified. Then the full-length cDNA was generated by PCR using primers specific to 5' and 3' ends of the gene by end-to-end PCR.

EXAMPLE II. CONSTRUCTION OF EXPRESSION VECTORS

The sequence was amplified from a genomic or cDNA library using primers specific to sequences upstream and downstream of the coding region. The expression vector was pMEN20 or pMEN65, which are both derived from pMON316 (Sanders et al, (1987) Nucleic Acids Research 15:1543-58) and contain the CaMV 35S promoter to express transgenes. To clone the sequence into the vector, both pMEN20 and the amplified DNA fragment were digested separately with SalI and NotI restriction enzymes at 37° C for 2 hours. The digestion products were subject to electrophoresis in a 0.8% agarose gel and visualized by ethidium bromide staining. The DNA fragments containing the sequence and the linearized plasmid were excised and purified by using a Qiaquick gel extraction kit (Qiagen, CA). The fragments of interest were ligated at a ratio of 3:1 (vector to insert). Ligation reactions using T4 DNA ligase (New England Biolabs, MA) were carried out at 16° C for 16 hours. The ligated DNAs were transformed into competent cells of the *E. coli* strain DH5alpha by using the heat shock method. The transformations were plated on LB plates containing 50 mg/l kanamycin (Sigma).

Individual colonies were grown overnight in five milliliters of LB broth containing 50 mg/l kanamycin at 37° C. Plasmid DNA was purified by using Qiaquick Mini Prep kits (Qiagen, CA).

EXAMPLE III. TRANSFORMATION OF *AGROBACTERIUM* WITH THE EXPRESSION VECTOR

After the plasmid vector containing the gene was constructed, the vector was used to transform *Agrobacterium tumefaciens* cells expressing the gene products. The stock of *Agrobacterium tumefaciens* cells for transformation were made as described by Nagel et al. (1990) FEMS Microbiol Letts. 67: 325-328. *Agrobacterium* strain ABI was grown in 250 ml LB medium (Sigma) overnight at 28°C with shaking until an absorbance (A_{600}) of 0.5 – 1.0 was

reached. Cells were harvested by centrifugation at 4,000 x g for 15 min at 4° C. Cells were then resuspended in 250 µl chilled buffer (1 mM HEPES, pH adjusted to 7.0 with KOH). Cells were centrifuged again as described above and resuspended in 125 µl chilled buffer. Cells were then centrifuged and resuspended two more times in the same HEPES buffer as described above at a volume of 100 µl and 750 µl, respectively. Resuspended cells were then distributed into 40 µl aliquots, quickly frozen in liquid nitrogen, and stored at -80° C.

Agrobacterium cells were transformed with plasmids prepared as described above following the protocol described by Nagel et al. For each DNA construct to be transformed, 50 – 100 ng DNA (generally resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was mixed with 40 µl of *Agrobacterium* cells. The DNA/cell mixture was then transferred to a chilled cuvette with a 2mm electrode gap and subject to a 2.5 kV charge dissipated at 25 µF and 200 µF using a Gene Pulser II apparatus (Bio-Rad). After electroporation, cells were immediately resuspended in 1.0 ml LB and allowed to recover without antibiotic selection for 2 – 4 hours at 28° C in a shaking incubator. After recovery, cells were plated onto selective medium of LB broth containing 100 µg/ml spectinomycin (Sigma) and incubated for 24-48 hours at 28° C. Single colonies were then picked and inoculated in fresh medium. The presence of the plasmid construct was verified by PCR amplification and sequence analysis.

EXAMPLE IV. TRANSFORMATION OF *ARABIDOPSIS* PLANTS WITH *AGROBACTERIUM TUMEFACIENS* WITH EXPRESSION VECTOR

After transformation of *Agrobacterium tumefaciens* with plasmid vectors containing the gene, single *Agrobacterium* colonies were identified, propagated, and used to transform *Arabidopsis* plants. Briefly, 500 ml cultures of LB medium containing 50 mg/l kanamycin were inoculated with the colonies and grown at 28° C with shaking for 2 days until an absorbance (A_{600}) of > 2.0 is reached. Cells were then harvested by centrifugation at 4,000 x g for 10 min, and resuspended in infiltration medium (1/2 X Murashige and Skoog salts (Sigma), 1 X Gamborg's B-5 vitamins (Sigma), 5.0% (w/v) sucrose (Sigma), 0.044 µM benzylamino purine (Sigma), 200 µl/L Silwet L-77 (Lehle Seeds) until an absorbance (A_{600}) of 0.8 was reached.

Prior to transformation, *Arabidopsis thaliana* seeds (ecotype Columbia) were sown at a density of ~10 plants per 4" pot onto Pro-Mix BX potting medium (Hummert International) covered with fiberglass mesh (18 mm X 16 mm). Plants were grown under continuous illumination (50-75 $\mu E/m^2/sec$) at 22-23° C with 65-70% relative humidity. After about 4 weeks, primary inflorescence stems (bolts) are cut off to encourage growth of multiple secondary bolts. After flowering of the mature secondary bolts, plants were prepared for transformation by removal of all siliques and opened flowers.

The pots were then immersed upside down in the mixture of *Agrobacterium* infiltration medium as described above for 30 sec, and placed on their sides to allow draining into a 1' x 2' flat surface covered with plastic wrap. After 24 h, the plastic wrap was removed and pots are turned upright. The immersion procedure was repeated one week later, for a total of two immersions per pot. Seeds were then collected from each transformation pot and analyzed following the protocol described below.

EXAMPLE V. IDENTIFICATION OF ARABIDOPSIS PRIMARY TRANSFORMANTS

Seeds collected from the transformation pots were sterilized essentially as follows. Seeds were dispersed into in a solution containing 0.1% (v/v) Triton X-100 (Sigma) and sterile H_2O and washed by shaking the suspension for 20 min. The wash solution was then drained and replaced with fresh wash solution to wash the seeds for 20 min with shaking. After removal of the second wash solution, a solution containing 0.1% (v/v) Triton X-100 and 70% ethanol (Equistar) was added to the seeds and the suspension was shaken for 5 min. After removal of the ethanol/detergent solution, a solution containing 0.1% (v/v) Triton X-100 and 30% (v/v) bleach (Clorox) was added to the seeds, and the suspension was shaken for 10 min. After removal of the bleach/detergent solution, seeds were then washed five times in sterile distilled H_2O. The seeds were stored in the last wash water at 4° C for 2 days in the dark before being plated onto antibiotic selection medium (1 X Murashige and Skoog salts (pH adjusted to 5.7 with 1M KOH), 1 X Gamborg's B-5 vitamins, 0.9% phytagar (Life Technologies), and 50 mg/l kanamycin). Seeds were germinated under continuous illumination (50-75 $\mu E/m^2/sec$) at 22-23° C. After 7-10 days of growth under these conditions, kanamycin resistant primary transformants (T_1 generation) were visible and obtained. These seedlings were transferred first to fresh selection plates where the seedlings continued to grow for 3-5 more days, and then to soil (Pro-Mix BX potting medium).

Primary transformants were crossed and progeny seeds (T_2) collected; kanamycin resistant seedlings were selected and analyzed. The expression levels of the recombinant polynucleotides in the transformants varies from about a 5% expression level increase to a least a 100% expression level increase. Similar observations are made with respect to polypeptide level expression.

EXAMPLE VI. IDENTIFICATION OF ARABIDOPSIS PLANTS WITH TRANSCRIPTION FACTOR GENE KNOCKOUTS

The screening of insertion mutagenized *Arabidopsis* collections for null mutants in a known target gene was essentially as described in Krysan et al (1999) Plant Cell 11:2283-2290. Briefly, gene-specific primers, nested by 5-250 base pairs to each other, were designed from the 5' and 3' regions of a known target gene. Similarly, nested sets of primers were also created specific to each of the T-DNA or transposon ends (the "right" and "left" borders). All possible combinations of gene specific and T-DNA/transposon primers were used to detect by PCR an insertion event within or close to the target gene. The amplified DNA fragments were then sequenced which allows the precise determination of the T-DNA/transposon insertion point relative to the target gene. Insertion events within the coding or intervening sequence of the genes were deconvoluted from a pool comprising a plurality of insertion events to a single unique mutant plant for functional characterization. The method is described in more detail in Yu and Adam, US Application Serial No. 09/177,733 filed October 23, 1998.

EXAMPLE VII. IDENTIFICATION OF MODIFIED BIOCHEMICAL CHARACTERISTICS PHENOTYPE IN OVEREXPRESSOR OR GENE KNOCKOUT PLANTS

Experiments were performed to identify those transformants or knockouts that exhibited modified biochemical characteristics. Among the biochemicals that were assayed were insoluble sugars, such as arabinose, fucose, galactose, mannose, rhamnose or xylose or the like; prenyl lipids, such as lutein, beta-carotene, xanthophyll-1, xanthophyll-2, chlorophylls A or B, or alpha-, delta- or gamma-tocopherol or the like; fatty acids, such as 16:0 (palmitic acid), 16:1 (palmitoleic acid), 18:0 (stearic acid), 18:1 (oleic acid), 18:2 (linoleic acid), 20:0 , 18:3 (linolenic acid), 20:1 (eicosenoic acid), 20:2, 22:1 (erucic acid) or the like; waxes, such as by altering the levels of C29, C31, or C33 alkanes; sterols, such as brassicasterol, campesterol, stigmasterol, sitosterol or stigmastanol or the like, glucosinolates, protein or oil levels

Fatty acids were measured using two methods depending on whether the tissue was from leaves or seeds. For leaves, lipids were extracted and esterified with hot methanolic H2SO4 and partitioned into hexane from methanolic brine. For seed fatty acids, seeds were pulverized and extracted in methanol:heptane:toluene:2,2-dimethoxypropane:H2SO4 (39:34:20:5:2) for 90 minutes at 80°C. After cooling to room temperature the upper phase, containing the seed fatty acid esters, was subjected to GC analysis. Fatty acid esters from both seed and leaf tissues were analyzed with a Supelco SP-2330 column.

Glucosinolates were purified from seeds or leaves by first heating the tissue at 95°C for 10 minutes. Preheated ethanol:water (50:50) is and after heating at 95°C for a further 10 minutes, the extraction solvent is applied to a DEAE Sephadex column which had been previously equilibrated with 0.5 M pyridine acetate. Desulfoglucosinolates were eluted with 300 ul water and analyzed by reverse phase HPLC monitoring at 226 nm.

For wax alkanes, samples were extracted using an identical method as fatty acids and extracts were analyzed on a HP 5890 GC coupled with a 5973 MSD. Samples were chromatographed on a J&W DB35 mass spectrometer (J&W Scientific).

To measure prenyl lipids levels, seeds or leaves were pulverized with 1 to 2% pyrogallol as an antioxidant. For seeds, extracted samples were filtered and a portion removed for tocopherol and carotenoid/chlorophyll analysis by HPLC. The remaining material was saponified for sterol determination. For leaves, an aliquot was removed and diluted with methanol and chlorophyll A, chlorophyll B, and total carotenoids measured by spectrophotometry by determining absorbance at 665.2 nm, 652.5 nm, and 470 nm. An aliquot was removed for tocopherol and carotenoid/chlorophyll composition by HPLC using a Waters uBondapak C18 column (4.6 mm x 150 mm). The remaining methanolic solution was saponified with 10% KOH at 80°C for one hour. The samples were cooled and diluted with a mixture of methanol and water. A solution of 2% methylene chloride in hexane was mixed in and the samples were centrifuged. The aqueous methanol phase was again re-extracted 2% methylene chloride in hexane and, after centrifugation, the two upper phases were combined and evaporated. 2% methylene chloride in hexane was added to the tubes and the samples were then extracted with one ml of water. The upper phase was removed, dried, and resuspended in 400 ul of 2% methylene chloride in hexane and analyzed by gas chromatography using a 50 m DB-5ms (0.25 mm ID, 0.25 um phase, J&W Scientific).

Insoluble sugar levels were measured by the method essentially described by Reiter et al., Plant Journal 12:335-345. This method analyzes the neutral sugar composition of cell wall polymers found in *Arabidopsis* leaves. Soluble sugars were separated from sugar polymers by extracting leaves with hot 70% ethanol. The remaining residue containing the insoluble polysaccharides was then acid hydrolyzed with allose added as an internal standard. Sugar monomers generated by the hydrolysis were then reduced to the corresponding alditols by treatment with NaBH4, then were acetylated to generate the volatile alditol acetates which were then analyzed by GC-FID. Identity of the peaks was determined by comparing the retention times of known sugars converted to the corresponding alditol acetates with the retention times of peaks from wild-type plant extracts. Alditol acetates were analyzed on a Supelco SP-2330 capillary

column (30 m x 250 um x 0.2 um) using a temperature program beginning at 180° C for 2 minutes followed by an increase to 220° C in 4 minutes. After holding at 220° C for 10 minutes, the oven temperature is increased to 240° C in 2 minutes and held at this temperature for 10 minutes and brought back to room temperature.

To identify plants with alterations in total seed oil or protein content, 150mg of seeds from T2 progeny plants were subjected to analysis by Near Infrared Reflectance (NIR) using a Foss NirSystems Model 6500 with a spinning cup transport system.

Table 3 shows the phenotypes observed for particular overexpressor or knockout plants and provides the SEQ ID No., the internal reference code (GID), whether a knockout or overexpressor plant was analyzed and the observed phenotype.

Table 3

SEQ ID No.	GID	Knockout (KO) or overexpressor (OE)	Phenotype observed
1	G214	OE	Increase in leaf fatty acids, for example 100% increase in 18:0 fatty acid. Also up to 100% increase in leaf chlorophyll and 100% increase in leaf carotenoids
3	G231	OE	Up to 5% increase in leaf 18:3 fatty acid
5	G274	OE	Up to 50% increase in leaf arabinose
7	G307	OE	Altered in leaf insoluble sugars, for example up to 44% decrease in mannose.
9	G346	OE	Altered leaf fatty acids, for example 25% increase in 16:3 and altered insoluble sugars, for example up to 25% increase in fucose
11	G598	OE	Altered in insoluble sugars, for example up to 20% decrease in rhamnose and up to 10% increase in galactose
13	G605	OE	Altered in leaf fatty acids, for example up to 20% increase in 16:1 fatty acid.
15	G777	OE	Altered in insoluble sugars, for example up to 60% increase in leaf rhamnose
17	G869	OE	Alteration in leaf fatty acids eg up to 39% decrease in 16:0 fatty acid; up to 43% increase in fucose
19	G1133	OE	Up to 34% decrease in leaf lutein
21	G1266	OE	Alteration in leaf fatty acids, for example up to 50% increase in 18:0 fatty acid. Alterations in leaf insoluble sugars, for example a 45% decrease in rhamnose
23	G1324	OE	Up to 65% decrease in leaf lutein and up to 84% increase in leaf xanthophyll

25	G1337	OE	Alteration in leaf fatty acids, for example up to 28% increase in 18:1 fatty acid
27	G975	OE	Up to 13-fold increase in wax in leaves

For a particular overexpressor that shows a less beneficial biochemical characteristic, it may be more useful to select a plant with a decreased expression of the particular transcription factor. For a particular knockout that shows a less beneficial biochemical characteristic, it may be more useful to select a plant with an increased expression of the particular transcription factor.

EXAMPLE VIII. IDENTIFICATION OF HOMOLOGOUS SEQUENCES

Homologous sequences from *Arabidopsis* and plant species other than *Arabidopsis* were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul et al. (1990) J. Mol. Biol. 215:403-410; and Altschul et al. (1997) Nucl. Acid Res. 25: 3389-3402). The tblastx sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff, S. and Henikoff, J. G. (1992) Proc. Natl. Acad. Sci. USA 89: 10915-10919).

Identified *Arabidopsis* homologous sequences are provided in Figure 2 and included in the Sequence Listing. The percent sequence identity among these sequences is as low as 47% sequence identity. Additionally, the entire NCBI GenBank database was filtered for sequences from all plants except *Arabidopsis thaliana* by selecting all entries in the NCBI GenBank database associated with NCBI taxonomic ID 33090 (Viridiplantae; all plants) and excluding entries associated with taxonomic ID 3701 (*Arabidopsis thaliana*). These sequences were compared to sequences representing genes of SEQ IDs Nos. 1-54 on 9/26/2000 using the Washington University TBLASTX algorithm (version 2.0a19MP). For each gene of SEQ IDs Nos. 1-54, individual comparisons were ordered by probability score (P-value), where the score reflects the probability that a particular alignment occurred by chance. For example, a score of 3.6e-40 is 3.6×10^{-40}. For up to ten species, the gene with the lowest P-value (and therefore the most likely homolog) is listed in Figure 3.

In addition to P-values, comparisons were also scored by percentage identity. Percentage identity reflects the degree to which two segments of DNA or protein are identical over a particular length. The ranges of percent identity between the non-Arabidopsis genes shown in Figure 3 and the Arabidopsis genes in the sequence listing are: SEQ ID No. 1: 38%-89%; SEQ ID No. 3: 64%-88%; SEQ ID No. 5: 44%-84%; SEQ ID No. 7: 35%-86%; SEQ ID No. 9: 43%-77%; SEQ ID No. 11: 43%-85%; SEQ ID No. 13: 41%-76%; SEQ ID No. 15: 34%-63%; SEQ ID No. 17: 31%-68%; SEQ ID No. 19: 26%-44%; SEQ ID No. 21: 52%-70%; SEQ ID No. 23: 37%-

93%; SEQ ID No. 25: 37%-58%; SEQ ID No. 27: 48%-92%; SEQ ID No. 29: 42%-88%; SEQ ID No. 31: 47%-90%; SEQ ID No. 33: 45%-69%; SEQ ID No. 35: 42%-94%; SEQ ID No. 37: 38%-85%; SEQ ID No. 39: 49%-93%; SEQ ID No. 41: 36%-64%; and SEQ ID No. 43: 36%-70%.

The polynucleotides and polypeptides in the Sequence Listing and the identified homologous sequences may be stored in a computer system and have associated or linked with the sequences a function, such as that the polynucleotides and polypeptides are useful for modifying the biochemical characteristics of a plant.

All references, publications, patents and other documents herein are incorporated by reference in their entirety for all purposes. Although the invention has been described with reference to the embodiments and examples above, it should be understood that various modifications can be made without departing from the spirit of the invention.

What is claimed is:

1. A transgenic plant with a modified biochemical characteristic, which plant comprises a recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID Nos. 2N, where N=1-22, or a complementary nucleotide sequence thereof;

(b) a nucleotide sequence encoding a polypeptide comprising a conservatively substituted variant of a polypeptide of (a);

(c) a nucleotide sequence comprising a sequence selected from those of SEQ ID Nos. 2N-1, where N=1-22, or a complementary nucleotide sequence thereof;

(d) a nucleotide sequence comprising silent substitutions in a nucleotide sequence of (c);

(e) a nucleotide sequence which hybridizes under stringent conditions to a nucleotide sequence of one or more of: (a), (b), (c), or (d);

(f) a nucleotide sequence comprising at least 15 consecutive nucleotides of a sequence of any of (a)-(e);

(g) a nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide that modifies a plant's biochemical characteristic;

(h) a nucleotide sequence having at least 31% sequence identity to a nucleotide sequence of any of (a)-(g);

(i) a nucleotide sequence having at least 60% identity sequence identity to a nucleotide sequence of any of (a)-(g);

(j) a nucleotide sequence which encodes a polypeptide having at least 31% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-22;

(k) a nucleotide sequence which encodes a polypeptide having at least 60% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-22; and

(l) a nucleotide sequence which encodes a polypeptide having at least 65% sequence identity to a conserved domain of a polypeptide of SEQ ID Nos. 2N, where N=1-22.

2. The transgenic plant of claim 1, further comprising a constitutive, inducible, or tissue-active promoter operably linked to said nucleotide sequence.

3. The transgenic plant of claim 1, wherein the plant is selected from the group consisting of: soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf,

banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, and vegetable brassicas.

4. An isolated or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID Nos. 2N, where N=1-22, or a complementary nucleotide sequence thereof;

(b) a nucleotide sequence encoding a polypeptide comprising a conservatively substituted variant of a polypeptide of (a);

(c) a nucleotide sequence comprising a sequence selected from those of SEQ ID Nos. 2N-1, where N=1-22, or a complementary nucleotide sequence thereof;

(d) a nucleotide sequence comprising silent substitutions in a nucleotide sequence of (c);

(e) a nucleotide sequence which hybridizes under stringent conditions to a nucleotide sequence of one or more of: (a), (b), (c), or (d);

(f) a nucleotide sequence comprising at least 15 consecutive nucleotides of a sequence of any of (a)-(e);

(g) a nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide that modifies a plant's biochemical characteristic;

(h) a nucleotide sequence having at least 31% sequence identity to a nucleotide sequence of any of (a)-(g);

(i) a nucleotide sequence having at least 60% identity sequence identity to a nucleotide sequence of any of (a)-(g);

(j) a nucleotide sequence which encodes a polypeptide having at least 31% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-22;

(k) a nucleotide sequence which encodes a polypeptide having at least 60% identity sequence identity to a polypeptide of SEQ ID Nos. 2N, where N=1-22; and

(l) a nucleotide sequence which encodes a conserved domain of a polypeptide having at least 65% sequence identity to a conserved domain of a polypeptide of SEQ ID Nos. 2N, where N=1-22.

5. The isolated or recombinant polynucleotide of claim 4, further comprising a constitutive, inducible, or tissue-active promoter operably linked to the nucleotide sequence.

6. A cloning or expression vector comprising the isolated or recombinant polynucleotide of claim 4.

7. A cell comprising the cloning or expression vector of claim 6.

8. A transgenic plant comprising the isolated or recombinant polynucleotide of claim 4.

9. A composition produced by one or more of:
(a) incubating one or more polynucleotide of claim 4 with a nuclease;
(b) incubating one or more polynucleotide of claim 4 with a restriction enzyme;
(c) incubating one or more polynucleotide of claim 4 with a polymerase;
(d) incubating one or more polynucleotide of claim 4 with a polymerase and a primer;
(e) incubating one or more polynucleotide of claim 4 with a cloning vector, or
(f) incubating one or more polynucleotide of claim 4 with a cell.

10. A composition comprising two or more different polynucleotides of claim 4.

11. An isolated or recombinant polypeptide comprising a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotide of claim 4.

12. A plant ectopically expressing an isolated polypeptide of claim 11.

13. A method for producing a plant having a modified biochemical characteristic, the method comprising altering the expression of the isolated or recombinant polynucleotide of claim 4 or the expression levels or activity of a polypeptide of claim 11 in a plant, thereby producing a modified plant, and selecting the modified plant for a modified biochemical characteristic thereby providing the modified plant with a modified biochemical characteristic.

14. The method of claim 13, wherein the polynucleotide is a polynucleotide of claim 4.

15. A method of identifying a factor that is modulated by or interacts with a polypeptide encoded by a polynucleotide of claim 4, the method comprising:

(a) expressing a polypeptide encoded by the polynucleotide in a plant; and

(b) identifying at least one factor that is modulated by or interacts with the polypeptide.

16. The method of claim 15, wherein the identifying is performed by detecting binding by the polypeptide to a promoter sequence, or detecting interactions between an additional protein and the polypeptide in a yeast two hybrid system.

17. The method of claim 15, wherein the identifying is performed by detecting expression of a factor by hybridization to a microarray, subtractive hybridization or differential display.

18. A method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide of interest, the method comprising:

(a) placing the molecule in contact with a plant comprising the polynucleotide or polypeptide encoded by the polynucleotide of claim 4; and,

(b) monitoring one or more of:

(i) expression level of the polynucleotide in the plant;

(ii) expression level of the polypeptide in the plant;

(iii) modulation of an activity of the polypeptide in the plant; or

(iv) modulation of an activity of the polynucleotide in the plant.

19. An integrated system, computer or computer readable medium comprising one or more character strings corresponding to a polynucleotide of claim 4, or to a polypeptide encoded by the polynucleotide.

20. The integrated system, computer or computer readable medium of claim 19, further comprising a link between said one or more sequence strings to a modified plant biochemical characteristics phenotype.

21. A method of identifying a sequence similar or homologous to one or more polynucleotides of claim 4, or one or more polypeptides encoded by the polynucleotides, the method comprising:

(a) providing a sequence database; and,

(b) querying the sequence database with one or more target sequences corresponding to the one or more polynucleotides or to the one or more polypeptides to identify one or more sequence members of the database that display sequence similarity or homology to one or more of the one or more target sequences.

22. The method of claim 21, wherein the querying comprises aligning one or more of the target sequences with one or more of the one or more sequence members in the sequence database.

23. The method of claim 21, wherein the querying comprises identifying one or more of the one or more sequence members of the database that meet a user-selected identity criteria with one or more of the target sequences.

24. The method of claim 21, further comprising linking the one or more of the polynucleotides of claim 4, or encoded polypeptides, to a modified plant biochemical characteristics phenotype.

25. A plant comprising altered expression levels of an isolated or recombinant polynucleotide of claim 4.

26. A plant comprising altered expression levels or the activity of an isolated or recombinant polypeptide of claim 11.

27. A plant lacking a nucleotide sequence encoding a polynucleotide of claim 11.

Figure 1

SEQ ID No.	GID	cDNA or protein	conserved domain
1	G214	cDNA	
2	G214	protein	22-71
3	G231	cDNA	
4	G231	protein	14-118
5	G274	cDNA	
6	G274	protein	108-572
7	G307	cDNA	
8	G307	protein	323-339
9	G346	cDNA	
10	G346	protein	196-221
11	G598	cDNA	
12	G598	protein	205-263
13	G605	cDNA	
14	G605	protein	132-143
15	G777	cDNA	
16	G777	protein	47-101
17	G869	cDNA	
18	G869	protein	109-177
19	G1133	cDNA	
20	G1133	protein	256-326
21	G1266	cDNA	
22	G1266	protein	79-147
23	G1324	cDNA	
24	G1324	protein	20-118
25	G1337	cDNA	
26	G1337	protein	9-75
27	G975	cDNA	
28	G975	protein	4-71

Figure 2

SEQ ID No.	GID	homolog	cDNA or protein	conserved domain
29	G680	homolog of G214	cDNA	
30	G680	homolog of G214	protein	24-70
31	G883	homolog of G274	cDNA	
32	G883	homolog of G274	protein	245-302
33	G1855	homolog of G274	cDNA	
34	G1855	homolog of G274	protein	entire protein
35	G1190	homolog of G274	cDNA	
36	G1190	homolog of G274	protein	entire protein
37	G308	homolog of G307	cDNA	
38	G308	homolog of G307	protein	270-274
39	G1944	homolog of G605	cDNA	
40	G1944	homolog of G605	protein	87-100
41	G326	homolog of G1337	cDNA	
42	G326	homolog of G1337	protein	11-94, 354-400
43	G1387	homolog of G975	cDNA	
44	G1387	homolog of G975	protein	4-71

Figure 3A

SEQ ID No.	GID	Genbank NID	P-value	Species
1	G214	8170933	8.80E-35	Lycopersicon esculentum
1	G214	9205339	1.20E-27	Glycine max
1	G214	8577344	1.80E-23	Zea mays
1	G214	9119112	2.40E-18	Medicago truncatula
1	G214	7660673	4.80E-15	Sorghum bicolor
1	G214	8213273	4.40E-14	Oryza sativa
1	G214	3325786	4.70E-10	Gossypium hirsutum
1	G214	9435251	1.50E-09	Hordeum vulgare
1	G214	9411569	6.80E-09	Triticum aestivum
1	G214	7614730	3.00E-07	Lotus japonicus
3	G231	6651291	7.80E-71	Pimpinella brachycarpa
3	G231	1430845	1.90E-62	Lycopersicon esculentum
3	G231	5268844	1.40E-61	Zea mays
3	G231	7561750	3.90E-60	Medicago truncatula
3	G231	1945282	3.30E-59	Oryza sativa
3	G231	22637	9.80E-49	Physcomitrella patens
3	G231	437326	2.00E-48	Gossypium hirsutum
3	G231	20562	3.40E-48	Petunia x hybrida
3	G231	4886263	5.00E-48	Antirrhinum majus
3	G231	8379692	1.50E-47	Gossypium arboreum
5	G274	6752887	1.70E-231	Malus domestica
5	G274	5734616	1.20E-140	Oryza sativa
5	G274	8996178	5.40E-96	Suaeda maritima subsp. salsa
5	G274	6654657	1.50E-89	Medicago truncatula
5	G274	8105703	2.30E-88	Lycopersicon esculentum
5	G274	7625402	4.00E-87	Gossypium arboreum
5	G274	7588836	2.10E-82	Glycine max
5	G274	5045979	1.30E-76	Gossypium hirsutum
5	G274	7324635	1.90E-71	Lycopersicon pennellii
5	G274	8903627	3.60E-63	Hordeum vulgare
7	G307	5640156	3.80E-151	Triticum aestivum
7	G307	5640154	1.00E-101	Zea mays
7	G307	6970471	1.70E-97	Oryza sativa
7	G307	7718432	4.00E-82	Medicago truncatula
7	G307	8330344	7.90E-78	Mesembryanthemum crystallinum
7	G307	5047560	1.00E-72	Gossypium hirsutum
7	G307	7588689	2.70E-69	Glycine max
7	G307	7623983	2.20E-64	Gossypium arboreum
7	G307	7780253	9.30E-59	Lotus japonicus
7	G307	6733213	1.90E-51	Lycopersicon esculentum
9	G346	4387642	5.90E-28	Lycopersicon esculentum
9	G346	7627902	1.50E-27	Gossypium arboreum
9	G346	8335147	6.40E-27	Oryza sativa
9	G346	8529362	9.10E-27	Medicago truncatula
9	G346	403305	2.30E-26	Nicotiana tabacum
9	G346	9299618	2.50E-26	Sorghum bicolor
9	G346	5056246	7.80E-26	Brassica rapa subsp. pekinensis
9	G346	6827291	6.80E-25	Zea mays
9	G346	6567406	1.90E-24	Glycine max
9	G346	9425896	1.20E-21	Triticum turgidum subsp. durum
11	G598	8102670	1.30E-43	Zea mays
11	G598	4382198	9.80E-42	Lycopersicon esculentum

Figure 3B

SEQ ID No.	GID	Genbank NID	P-value	Species
11	G598	7553316	8.00E-38	Sorghum bicolor
11	G598	9445834	3.10E-36	Triticum aestivum
11	G598	7332502	8.80E-30	Oryza sativa
11	G598	9056816	1.70E-17	Medicago truncatula
11	G598	6644720	5.20E-15	Mesembryanthemum crystallinum
11	G598	3853398	2.20E-14	Populus tremula x Populus tremuloides
11	G598	9419408	6.80E-09	Hordeum vulgare
11	G598	6848223	1.40E-06	Glycine max
13	G605	7624850	4.40E-49	Gossypium arboreum
13	G605	9204125	6.50E-46	Glycine max
13	G605	2213533	5.50E-33	Pisum sativum
13	G605	7009437	1.40E-28	Zea mays
13	G605	8104258	3.50E-28	Lycopersicon esculentum
13	G605	7536402	4.10E-28	Sorghum bicolor
13	G605	3107210	1.60E-22	Oryza sativa
13	G605	7784135	9.20E-20	Lotus japonicus
13	G605	4165182	8.30E-18	Antirrhinum majus
13	G605	6555294	8.10E-17	Pinus taeda
15	G777	8172576	3.10E-29	Medicago truncatula
15	G777	8331320	4.60E-17	Mesembryanthemum crystallinum
15	G777	8106138	3.00E-16	Lycopersicon esculentum
15	G777	5046832	1.20E-14	Gossypium hirsutum
15	G777	6918785	1.70E-13	Zea mays
15	G777	5666914	1.30E-07	Glycine max
15	G777	8856987	0.98	Oryza sativa
15	G777	8404755	1	Hordeum vulgare
17	G869	2213784	1.30E-19	Lycopersicon esculentum
17	G869	3065894	7.30E-19	Nicotiana tabacum
17	G869	8570080	4.20E-18	Oryza sativa
17	G869	7560260	1.50E-17	Medicago truncatula
17	G869	7534890	5.20E-14	Sorghum bicolor
17	G869	6455322	1.10E-13	Glycine max
17	G869	9362061	2.70E-13	Triticum aestivum
17	G869	7788764	5.70E-13	Lotus japonicus
17	G869	7624302	2.50E-12	Gossypium arboreum
17	G869	3858036	2.80E-12	Populus balsamifera subsp. trichocarpa
19	G1133	8070726	1.30E-16	Solanum tuberosum
19	G1133	6848196	1.60E-16	Glycine max
19	G1133	7570922	3.60E-13	Medicago truncatula
19	G1133	9434859	1.90E-12	Lycopersicon esculentum
19	G1133	5704484	0.005	Oryza sativa
19	G1133	902661	0.0081	Hordeum vulgare
19	G1133	8666194	0.0086	Pinus taeda
19	G1133	5725018	0.14	Brassica rapa subsp. pekinensis
19	G1133	7501051	0.64	Gossypium arboreum
19	G1133	7747388	0.98	Lotus japonicus
21	G1266	1732405	1.50E-50	Nicotiana tabacum
21	G1266	7145976	2.50E-38	Glycine max
21	G1266	3326366	1.00E-37	Gossypium hirsutum
21	G1266	5762854	6.90E-37	Lotus japonicus
21	G1266	7560749	9.10E-34	Medicago truncatula
21	G1266	7934594	6.60E-33	Euphorbia esula
21	G1266	9431305	2.10E-28	Lycopersicon esculentum

Figure 3C

SEQ ID No.	GID	Genbank NID	P-value	Species
21	G1266	7528275	5.40E-21	Mesembryanthemum crystallinum
21	G1266	6478844	4.10E-20	Matricaria chamomilla
21	G1266	7627061	4.20E-20	Gossypium arboreum
23	G1324	2921337	2.30E-54	Gossypium hirsutum
23	G1324	5891412	3.50E-52	Lycopersicon esculentum
23	G1324	8528843	7.20E-50	Medicago truncatula
23	G1324	1002797	5.40E-49	Craterostigma plantagineum
23	G1324	5666961	3.90E-44	Glycine max
23	G1324	7244640	1.70E-42	Mentha x piperita
23	G1324	1841474	3.00E-42	Pisum sativum
23	G1324	4979554	1.30E-39	Oryza sativa
23	G1324	9363368	3.00E-32	Triticum aestivum
23	G1324	9296080	3.50E-32	Sorghum bicolor
25	G1337	7410432	2.60E-41	Lycopersicon esculentum
25	G1337	3618319	1.10E-32	Oryza sativa
25	G1337	7571599	1.00E-28	Medicago truncatula
25	G1337	7685955	5.10E-27	Glycine max
25	G1337	7323708	2.60E-25	Lycopersicon hirsutum
25	G1337	4091805	1.00E-18	Malus domestica
25	G1337	6917805	4.80E-18	Lycopersicon pennellii
25	G1337	3341722	1.60E-17	Raphanus sativus
25	G1337	2303680	4.50E-17	Brassica napus
25	G1337	4557092	9.10E-17	Pinus radiata
27	G975	8103850	8.50E-46	Lycopersicon esculentum
27	G975	7590215	1.50E-45	Glycine max
27	G975	5056299	2.20E-34	Brassica rapa subsp. pekinensis
27	G975	9278522	1.80E-26	Lotus japonicus
27	G975	1128767	2.70E-18	Brassica rapa
27	G975	5859978	5.50E-18	Pinus taeda
27	G975	9427282	2.40E-15	Triticum aestivum
27	G975	19506	4.70E-14	Lupinus polyphyllus
27	G975	6799584	5.30E-14	Medicago truncatula
27	G975	7324705	1.70E-12	Lycopersicon pennellii
29	G680	9258166	5.70E-36	Glycine max
29	G680	9255178	3.00E-29	Zea mays
29	G680	5274804	1.20E-27	Lycopersicon esculentum
29	G680	4974199	3.00E-22	Oryza sativa
29	G680	3325786	2.10E-21	Gossypium hirsutum
29	G680	9119112	1.30E-18	Medicago truncatula
29	G680	7660673	3.20E-17	Sorghum bicolor
29	G680	7243970	6.10E-16	Mentha x piperita
29	G680	3858093	2.10E-10	Populus balsamifera subsp. trichocarpa
29	G680	8845091	3.70E-10	Triticum aestivum
31	G883	4760595	2.40E-84	Nicotiana tabacum
31	G883	4894962	3.50E-45	Avena sativa
31	G883	6719425	1.70E-36	Glycine max
31	G883	5273248	2.80E-35	Lycopersicon esculentum
31	G883	9302479	3.00E-34	Sorghum bicolor
31	G883	6799932	1.40E-31	Medicago truncatula
31	G883	5456433	4.30E-31	Zea mays
31	G883	8706346	1.40E-30	Hordeum vulgare
31	G883	8404566	2.70E-30	Oryza sativa
31	G883	1432055	2.00E-27	Petroselinum crispum

Figure 3D

SEQ ID No.	GID	Genbank NID	P-value	Species
33	G1855	6752887	4.80E-181	Malus domestica
33	G1855	5734616	7.60E-154	Oryza sativa
33	G1855	4384552	3.80E-80	Lycopersicon esculentum
33	G1855	8996178	1.80E-78	Suaeda maritima subsp. salsa
33	G1855	7625402	1.60E-77	Gossypium arboreum
33	G1855	8903627	3.80E-74	Hordeum vulgare
33	G1855	6654657	2.20E-70	Medicago truncatula
33	G1855	8090141	4.50E-64	Sorghum bicolor
33	G1855	9028645	6.30E-64	Zea mays
33	G1855	7588836	6.70E-62	Glycine max
35	G1190	6752887	7.00E-111	Malus domestica
35	G1190	5734616	1.20E-98	Oryza sativa
35	G1190	7569650	2.60E-92	Medicago truncatula
35	G1190	4380101	5.50E-88	Lycopersicon esculentum
35	G1190	6567183	5.20E-81	Glycine max
35	G1190	8901706	7.40E-80	Hordeum vulgare
35	G1190	8070121	1.70E-76	Solanum tuberosum
35	G1190	8666639	5.50E-75	Pinus taeda
35	G1190	8088688	3.40E-72	Sorghum bicolor
35	G1190	6020980	6.50E-67	Zea mays
37	G308	5640156	3.50E-162	Triticum aestivum
37	G308	5640154	2.30E-134	Zea mays
37	G308	6970471	4.20E-120	Oryza sativa
37	G308	7718432	8.70E-80	Medicago truncatula
37	G308	8330344	3.90E-76	Mesembryanthemum crystallinum
37	G308	5047560	1.50E-71	Gossypium hirsutum
37	G308	7588689	1.90E-68	Glycine max
37	G308	7623983	2.90E-62	Gossypium arboreum
37	G308	7780253	1.10E-57	Lotus japonicus
37	G308	6733213	3.70E-48	Lycopersicon esculentum
39	G1944	9204125	5.50E-52	Glycine max
39	G1944	7624850	6.60E-45	Gossypium arboreum
39	G1944	7784135	7.20E-32	Lotus japonicus
39	G1944	9280727	2.60E-29	Oryza sativa
39	G1944	7009437	1.30E-28	Zea mays
39	G1944	7536402	1.30E-28	Sorghum bicolor
39	G1944	8104258	6.50E-27	Lycopersicon esculentum
39	G1944	2213533	3.50E-23	Pisum sativum
39	G1944	4165182	7.10E-17	Antirrhinum majus
39	G1944	6555294	2.90E-16	Pinus taeda
41	G326	7410432	1.10E-37	Lycopersicon esculentum
41	G326	3618319	2.90E-32	Oryza sativa
41	G326	7571599	4.90E-30	Medicago truncatula
41	G326	7232283	6.30E-28	Glycine max
41	G326	7323708	6.00E-27	Lycopersicon hirsutum
41	G326	4091805	2.30E-19	Malus domestica
41	G326	6917805	6.50E-19	Lycopersicon pennellii
41	G326	3341722	2.50E-18	Raphanus sativus
41	G326	4557092	7.50E-18	Pinus radiata
41	G326	2303680	4.70E-17	Brassica napus
43	G1387	8285738	1.40E-46	Glycine max
43	G1387	8103850	5.20E-46	Lycopersicon esculentum
43	G1387	5056299	1.10E-20	Brassica rapa subsp. pekinensis

Figure 3E

SEQ ID No.	GID	Genbank NID	P-value	Species
43	G1387	9278522	1.50E-18	Lotus japonicus
43	G1387	5859978	2.00E-15	Pinus taeda
43	G1387	7766740	4.70E-14	Medicago truncatula
43	G1387	9427282	1.40E-12	Triticum aestivum
43	G1387	3857766	3.40E-12	Populus balsamifera subsp. trichocarpa
43	G1387	19506	4.60E-12	Lupinus polyphyllus
43	G1387	7273843	2.20E-11	Oryza sativa

MBI-20 Sequence Listing.ST25

SEQUENCE LISTING

<110> Creelman, Robert
Yu, Guo-Liang
Adam, Luc
Riechmann, Jose Luis
Heard, Jacqueline
Samaha, Raymond
Pilgrim, Marsha
Pineda, Omaira
Jiang, Cai-Zhong

<120> Plant Biochemistry-Related Genes

<130> MBI-0020

<150> 60/164,132
<151> 1999-11-17

<150> 60/197,899
<151> 2000-04-17

<150> Plant Trait Modification III
<151> 2000-08-22

<160> 44

<170> PatentIn version 3.0

<210> 1
<211> 2240
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (238)..(2064)
<223> G214

<400> 1
tgagatttct ccatttccgt agcttctggt ctcttttctt tgtttcattg atcaaaagca 60

aatcacttct tcttcttctt cttctcgatt tcttactgtt ttcttatcca acgaaatctg 120

gaattaaaaa tggaatcttt atcgaatcca agctgatttt gtttctttca ttgaatcatc 180

tctctaaagt ggaattttgt aaagagaaga tctgaagttg tgtagaggag cttagtg 237

atg gag aca aat tcg tct gga gaa gat ctg gtt att aag act cgg aag 285
Met Glu Thr Asn Ser Ser Gly Glu Asp Leu Val Ile Lys Thr Arg Lys
1 5 10 15

cca tat acg ata aca aag caa cgt gaa agg tgg act gag gaa gaa cat 333
Pro Tyr Thr Ile Thr Lys Gln Arg Glu Arg Trp Thr Glu Glu Glu His
20 25 30

aat aga ttc att gaa gct ttg agg ctt tat ggt aga gca tgg cag aag 381
Asn Arg Phe Ile Glu Ala Leu Arg Leu Tyr Gly Arg Ala Trp Gln Lys
35 40 45

att gaa gaa cat gta gca aca aaa act gct gtc cag ata aga agt cac 429
Ile Glu Glu His Val Ala Thr Lys Thr Ala Val Gln Ile Arg Ser His
50 55 60

gct cag aaa ttt ttc tcc aag gta gag aaa gag gct gaa gct aaa ggt 477
Ala Gln Lys Phe Phe Ser Lys Val Glu Lys Glu Ala Glu Ala Lys Gly
65 70 75 80

gta gct atg ggt caa gcg cta gac ata gct att cct cct cca cgg cct 525
Val Ala Met Gly Gln Ala Leu Asp Ile Ala Ile Pro Pro Pro Arg Pro
85 90 95

aag cgt aaa cca aac aat cct tat cct cga aag acg gga agt gga acg 573
Lys Arg Lys Pro Asn Asn Pro Tyr Pro Arg Lys Thr Gly Ser Gly Thr

```
          100                 105                 110

atc ctt atg tca aaa acg ggt gtg aat gat gga aaa gag tcc ctt gga      621
Ile Leu Met Ser Lys Thr Gly Val Asn Asp Gly Lys Glu Ser Leu Gly
        115                 120                 125

tca gaa aaa gtg tcg cat cct gag atg gcc aat gaa gat cga caa caa      669
Ser Glu Lys Val Ser His Pro Glu Met Ala Asn Glu Asp Arg Gln Gln
    130                 135                 140

tca aag cct gaa gag aaa act ctg cag gaa gac aac tgt tca gat tgt      717
Ser Lys Pro Glu Glu Lys Thr Leu Gln Glu Asp Asn Cys Ser Asp Cys
145                 150                 155                 160

ttc act cat cag tat ctc tct gct gca tcc tcc atg aat aaa agt tgt      765
Phe Thr His Gln Tyr Leu Ser Ala Ala Ser Ser Met Asn Lys Ser Cys
                165                 170                 175

ata gag aca tca aac gca agc act ttc cgc gag ttc ttg cct tca cgg      813
Ile Glu Thr Ser Asn Ala Ser Thr Phe Arg Glu Phe Leu Pro Ser Arg
            180                 185                 190

gaa gag gga agt cag aat aac agg gta aga aag gag tca aac tca gat      861
Glu Glu Gly Ser Gln Asn Asn Arg Val Arg Lys Glu Ser Asn Ser Asp
        195                 200                 205

ttg aat gca aaa tct ctg gaa aac ggt aat gag caa gga cct cag act      909
Leu Asn Ala Lys Ser Leu Glu Asn Gly Asn Glu Gln Gly Pro Gln Thr
    210                 215                 220

tat ccg atg cat atc cct gtg cta gtg cca ttg ggg agc tca ata aca      957
Tyr Pro Met His Ile Pro Val Leu Val Pro Leu Gly Ser Ser Ile Thr
225                 230                 235                 240

agt tct cta tca cat cct cct tca gag cca gat agt cat ccc cac aca     1005
Ser Ser Leu Ser His Pro Pro Ser Glu Pro Asp Ser His Pro His Thr
                245                 250                 255

gtt gca gga gat tat cag tcg ttt cct aat cat ata atg tca acc ctt     1053
Val Ala Gly Asp Tyr Gln Ser Phe Pro Asn His Ile Met Ser Thr Leu
            260                 265                 270

tta caa aca ccg gct ctt tat act gcc gca act ttc gcc tca tca ttt     1101
Leu Gln Thr Pro Ala Leu Tyr Thr Ala Ala Thr Phe Ala Ser Ser Phe
        275                 280                 285

tgg cct ccc gat tct agt ggt ggc tca cct gtt cca ggg aac tca cct     1149
Trp Pro Pro Asp Ser Ser Gly Gly Ser Pro Val Pro Gly Asn Ser Pro
    290                 295                 300

ccg aat ctg gct gcc atg gcc gca gcc act gtt gca gct gct agt gct     1197
Pro Asn Leu Ala Ala Met Ala Ala Ala Thr Val Ala Ala Ala Ser Ala
305                 310                 315                 320

tgg tgg gct gcc aat gga tta tta cct tta tgt gct cct ctt agt tca     1245
Trp Trp Ala Ala Asn Gly Leu Leu Pro Leu Cys Ala Pro Leu Ser Ser
                325                 330                 335

ggt ggt ttc act agt cat cct cca tct act ttt gga cca tca tgt gat     1293
Gly Gly Phe Thr Ser His Pro Pro Ser Thr Phe Gly Pro Ser Cys Asp
            340                 345                 350

gta gag tac aca aaa gca agc act tta caa cat ggt tct gtg cag agc     1341
Val Glu Tyr Thr Lys Ala Ser Thr Leu Gln His Gly Ser Val Gln Ser
        355                 360                 365

cga gag caa gaa cac tcc gag gca tca aag gct cga tct tca ctg gac     1389
Arg Glu Gln Glu His Ser Glu Ala Ser Lys Ala Arg Ser Ser Leu Asp
    370                 375                 380

tca gag gat gtt gaa aat aag agt aaa cca gtt tgt cat gag cag cct     1437
Ser Glu Asp Val Glu Asn Lys Ser Lys Pro Val Cys His Glu Gln Pro
385                 390                 395                 400

tct gca aca cct gag agt gat gca aag ggt tca gat gga gca gga gac     1485
```

MBI-20 Sequence Listing.ST25

```
Ser Ala Thr Pro Glu Ser Asp Ala Lys Gly Ser Asp Gly Ala Gly Asp
                405                 410                 415

aga aaa caa gtt gac cgg tcc tcg tgt ggc tca aac act ccg tcg agt     1533
Arg Lys Gln Val Asp Arg Ser Ser Cys Gly Ser Asn Thr Pro Ser Ser
            420                 425                 430

agt gat gat gtt gag gcg gat gca tca gaa agg caa gag gat ggc acc     1581
Ser Asp Asp Val Glu Ala Asp Ala Ser Glu Arg Gln Glu Asp Gly Thr
        435                 440                 445

aat ggt gag gtg aaa gaa acg aat gaa gac act aat aaa cct caa act     1629
Asn Gly Glu Val Lys Glu Thr Asn Glu Asp Thr Asn Lys Pro Gln Thr
    450                 455                 460

tca gag tcc aat gca cgc cgc agt aga atc agc tcc aat ata acc gat     1677
Ser Glu Ser Asn Ala Arg Arg Ser Arg Ile Ser Ser Asn Ile Thr Asp
465                 470                 475                 480

cca tgg aag tct gtg tct gac gag ggt cga att gcc ttc caa gct ctc     1725
Pro Trp Lys Ser Val Ser Asp Glu Gly Arg Ile Ala Phe Gln Ala Leu
                485                 490                 495

ttc tcc aga gag gta ttg ccg caa agt ttt aca tat cga gaa gaa cac     1773
Phe Ser Arg Glu Val Leu Pro Gln Ser Phe Thr Tyr Arg Glu Glu His
            500                 505                 510

aga gag gaa gaa caa caa caa caa gaa caa aga tat cca atg gca ctt     1821
Arg Glu Glu Glu Gln Gln Gln Gln Glu Gln Arg Tyr Pro Met Ala Leu
        515                 520                 525

gat ctt aac ttc aca gct cag tta aca cca gtt gat gat caa gag gag     1869
Asp Leu Asn Phe Thr Ala Gln Leu Thr Pro Val Asp Asp Gln Glu Glu
    530                 535                 540

aag aga aac aca gga ttt ctt gga atc gga tta gat gct tca aag cta     1917
Lys Arg Asn Thr Gly Phe Leu Gly Ile Gly Leu Asp Ala Ser Lys Leu
545                 550                 555                 560

atg agt aga gga aga aca ggt ttt aaa cca tac aaa aga tgt tcc atg     1965
Met Ser Arg Gly Arg Thr Gly Phe Lys Pro Tyr Lys Arg Cys Ser Met
                565                 570                 575

gaa gcc aaa gaa agt aga atc ctc aac aac aat cct atc att cat gtg     2013
Glu Ala Lys Glu Ser Arg Ile Leu Asn Asn Asn Pro Ile Ile His Val
            580                 585                 590

gaa cag aaa gat ccc aaa cgg atg cgg ttg gaa act caa gct tcc aca     2061
Glu Gln Lys Asp Pro Lys Arg Met Arg Leu Glu Thr Gln Ala Ser Thr
        595                 600                 605

tga gactctattt tcatctgatc tgttgtttgt actctgtttt taagttttca          2114

agaccactgc tacattttct ttttcttttg aggcctttgt atttgtttcc ttgtccatag   2174

tcttcctgta acatttgact ctgtattatt caacaaatca taaactgttt aatctttttt   2234

tttcca                                                             2240
```

<210> 2
<211> 608
<212> PRT
<213> Arabidopsis thaliana

<400> 2

```
Met Glu Thr Asn Ser Ser Gly Glu Asp Leu Val Ile Lys Thr Arg Lys
1               5                   10                  15

Pro Tyr Thr Ile Thr Lys Gln Arg Glu Arg Trp Thr Glu Glu Glu His
            20                  25                  30
```

MBI-20 Sequence Listing.ST25

```
Asn Arg Phe Ile Glu Ala Leu Arg Leu Tyr Gly Arg Ala Trp Gln Lys
        35                  40                  45

Ile Glu Glu His Val Ala Thr Lys Thr Ala Val Gln Ile Arg Ser His
    50                  55                  60

Ala Gln Lys Phe Phe Ser Lys Val Glu Lys Glu Ala Glu Ala Lys Gly
65                  70                  75                  80

Val Ala Met Gly Gln Ala Leu Asp Ile Ala Ile Pro Pro Pro Arg Pro
                85                  90                  95

Lys Arg Lys Pro Asn Asn Pro Tyr Pro Arg Lys Thr Gly Ser Gly Thr
            100                 105                 110

Ile Leu Met Ser Lys Thr Gly Val Asn Asp Gly Lys Glu Ser Leu Gly
        115                 120                 125

Ser Glu Lys Val Ser His Pro Glu Met Ala Asn Glu Asp Arg Gln Gln
    130                 135                 140

Ser Lys Pro Glu Glu Lys Thr Leu Gln Glu Asp Asn Cys Ser Asp Cys
145                 150                 155                 160

Phe Thr His Gln Tyr Leu Ser Ala Ala Ser Ser Met Asn Lys Ser Cys
                165                 170                 175

Ile Glu Thr Ser Asn Ala Ser Thr Phe Arg Glu Phe Leu Pro Ser Arg
            180                 185                 190

Glu Glu Gly Ser Gln Asn Asn Arg Val Arg Lys Glu Ser Asn Ser Asp
        195                 200                 205

Leu Asn Ala Lys Ser Leu Glu Asn Gly Asn Glu Gln Gly Pro Gln Thr
    210                 215                 220

Tyr Pro Met His Ile Pro Val Leu Val Pro Leu Gly Ser Ser Ile Thr
225                 230                 235                 240

Ser Ser Leu Ser His Pro Pro Ser Glu Pro Asp Ser His Pro His Thr
                245                 250                 255

Val Ala Gly Asp Tyr Gln Ser Phe Pro Asn His Ile Met Ser Thr Leu
            260                 265                 270

Leu Gln Thr Pro Ala Leu Tyr Thr Ala Ala Thr Phe Ala Ser Ser Phe
        275                 280                 285

Trp Pro Pro Asp Ser Ser Gly Gly Ser Pro Val Pro Gly Asn Ser Pro
    290                 295                 300

Pro Asn Leu Ala Ala Met Ala Ala Ala Thr Val Ala Ala Ala Ser Ala
305                 310                 315                 320

Trp Trp Ala Ala Asn Gly Leu Leu Pro Leu Cys Ala Pro Leu Ser Ser
                325                 330                 335
```

MBI-20 Sequence Listing.ST25

Gly Gly Phe Thr Ser His Pro Pro Ser Thr Phe Gly Pro Ser Cys Asp
340 345 350

Val Glu Tyr Thr Lys Ala Ser Thr Leu Gln His Gly Ser Val Gln Ser
355 360 365

Arg Glu Gln Glu His Ser Glu Ala Ser Lys Ala Arg Ser Ser Leu Asp
370 375 380

Ser Glu Asp Val Glu Asn Lys Ser Lys Pro Val Cys His Glu Gln Pro
385 390 395 400

Ser Ala Thr Pro Glu Ser Asp Ala Lys Gly Ser Asp Gly Ala Gly Asp
405 410 415

Arg Lys Gln Val Asp Arg Ser Ser Cys Gly Ser Asn Thr Pro Ser Ser
420 425 430

Ser Asp Asp Val Glu Ala Asp Ala Ser Glu Arg Gln Glu Asp Gly Thr
435 440 445

Asn Gly Glu Val Lys Glu Thr Asn Glu Asp Thr Asn Lys Pro Gln Thr
450 455 460

Ser Glu Ser Asn Ala Arg Arg Ser Arg Ile Ser Ser Asn Ile Thr Asp
465 470 475 480

Pro Trp Lys Ser Val Ser Asp Glu Gly Arg Ile Ala Phe Gln Ala Leu
485 490 495

Phe Ser Arg Glu Val Leu Pro Gln Ser Phe Thr Tyr Arg Glu Glu His
500 505 510

Arg Glu Glu Glu Gln Gln Gln Gln Glu Gln Arg Tyr Pro Met Ala Leu
515 520 525

Asp Leu Asn Phe Thr Ala Gln Leu Thr Pro Val Asp Asp Gln Glu Glu
530 535 540

Lys Arg Asn Thr Gly Phe Leu Gly Ile Gly Leu Asp Ala Ser Lys Leu
545 550 555 560

Met Ser Arg Gly Arg Thr Gly Phe Lys Pro Tyr Lys Arg Cys Ser Met
565 570 575

Glu Ala Lys Glu Ser Arg Ile Leu Asn Asn Asn Pro Ile Ile His Val
580 585 590

Glu Gln Lys Asp Pro Lys Arg Met Arg Leu Glu Thr Gln Ala Ser Thr
595 600 605

<210> 3
<211> 916
<212> DNA
<213> Arabidopsis thaliana

<220>

MBI-20 Sequence Listing.ST25

<221> CDS
<222> (88)..(888)
<223> G231

<400> 3

```
ttccatatct cttccatttc gctctctatt tcacatcccc atataacata atatacaatc      60

acacatatca tttctatata gtattta atg ggg aga cag cca tgc tgt gac aag     114
                               Met Gly Arg Gln Pro Cys Cys Asp Lys
                               1               5

cta ggg gtg aag aaa ggg ccg tgg acg gtg gag gaa gat aag aag ctt       162
Leu Gly Val Lys Lys Gly Pro Trp Thr Val Glu Glu Asp Lys Lys Leu
10                  15                  20                  25

ata aac ttc ata cta acc aat ggc cat tgt tgc tgg cgt gct ttg ccg       210
Ile Asn Phe Ile Leu Thr Asn Gly His Cys Cys Trp Arg Ala Leu Pro
                30                  35                  40

aag ctg gcc ggt ctc cgt cgc tgt gga aag agc tgc cgc ctc cgg tgg       258
Lys Leu Ala Gly Leu Arg Arg Cys Gly Lys Ser Cys Arg Leu Arg Trp
            45                  50                  55

act aac tat ctc cgg cct ggc tta aaa cga ggc ctt ctc tcg cat gat       306
Thr Asn Tyr Leu Arg Pro Gly Leu Lys Arg Gly Leu Leu Ser His Asp
        60                  65                  70

gaa gaa caa ctt gtc ata gat ctt cat gct aat ctc ggc aat aag tgg       354
Glu Glu Gln Leu Val Ile Asp Leu His Ala Asn Leu Gly Asn Lys Trp
    75                  80                  85

tct aag ata gct tca aga tta cct gga aga aca gat aac gaa ata aaa       402
Ser Lys Ile Ala Ser Arg Leu Pro Gly Arg Thr Asp Asn Glu Ile Lys
90                  95                  100                 105

aac cat tgg aat act cat atc aag aag aaa ctt ctt aag atg gga atc       450
Asn His Trp Asn Thr His Ile Lys Lys Lys Leu Leu Lys Met Gly Ile
                110                 115                 120

gat cct atg acc cat caa ccc cta aat caa gaa cct tct aat atc gat       498
Asp Pro Met Thr His Gln Pro Leu Asn Gln Glu Pro Ser Asn Ile Asp
            125                 130                 135

aat tcc aaa acc att ccg tcc aat cca gac gat gtc tca gtg gaa cca       546
Asn Ser Lys Thr Ile Pro Ser Asn Pro Asp Asp Val Ser Val Glu Pro
        140                 145                 150

aag aca act aac acg aaa tac gtg gag ata agt gtc acg aca aca gaa       594
Lys Thr Thr Asn Thr Lys Tyr Val Glu Ile Ser Val Thr Thr Thr Glu
    155                 160                 165

gaa gaa agt agt agc acg gtt act gat caa aac agt tcg atg gat aat       642
Glu Glu Ser Ser Ser Thr Val Thr Asp Gln Asn Ser Ser Met Asp Asn
170                 175                 180                 185

gaa aat cat cta att gac aac att tat gat gat gat gaa ttg ttt agt       690
Glu Asn His Leu Ile Asp Asn Ile Tyr Asp Asp Asp Glu Leu Phe Ser
                190                 195                 200

tac tta tgg tcc gac gaa act act aaa gat gag gcc tct tgg agt gat       738
Tyr Leu Trp Ser Asp Glu Thr Thr Lys Asp Glu Ala Ser Trp Ser Asp
            205                 210                 215

agt aac ttt ggt gtt ggt gga aca tta tat gac cac aat atc tcc ggc       786
Ser Asn Phe Gly Val Gly Gly Thr Leu Tyr Asp His Asn Ile Ser Gly
        220                 225                 230

gcc gat gca gat ttt ccg ata tgg tca ccg gaa aga atc aat gac gag       834
Ala Asp Ala Asp Phe Pro Ile Trp Ser Pro Glu Arg Ile Asn Asp Glu
    235                 240                 245

aag atg ttt ttg gat tat tgt caa gac ttt ggt gtt cat gat ttt ggg       882
Lys Met Phe Leu Asp Tyr Cys Gln Asp Phe Gly Val His Asp Phe Gly
250                 255                 260                 265
```

ttt tga ctgttcacca ttgacatatt ggcaacgc 916
Phe

<210> 4
<211> 266
<212> PRT
<213> Arabidopsis thaliana

<400> 4

Met Gly Arg Gln Pro Cys Cys Asp Lys Leu Gly Val Lys Lys Gly Pro
1 5 10 15

Trp Thr Val Glu Glu Asp Lys Lys Leu Ile Asn Phe Ile Leu Thr Asn
20 25 30

Gly His Cys Cys Trp Arg Ala Leu Pro Lys Leu Ala Gly Leu Arg Arg
35 40 45

Cys Gly Lys Ser Cys Arg Leu Arg Trp Thr Asn Tyr Leu Arg Pro Gly
50 55 60

Leu Lys Arg Gly Leu Leu Ser His Asp Glu Glu Gln Leu Val Ile Asp
65 70 75 80

Leu His Ala Asn Leu Gly Asn Lys Trp Ser Lys Ile Ala Ser Arg Leu
85 90 95

Pro Gly Arg Thr Asp Asn Glu Ile Lys Asn His Trp Asn Thr His Ile
100 105 110

Lys Lys Lys Leu Leu Lys Met Gly Ile Asp Pro Met Thr His Gln Pro
115 120 125

Leu Asn Gln Glu Pro Ser Asn Ile Asp Asn Ser Lys Thr Ile Pro Ser
130 135 140

Asn Pro Asp Asp Val Ser Val Glu Pro Lys Thr Thr Asn Thr Lys Tyr
145 150 155 160

Val Glu Ile Ser Val Thr Thr Thr Glu Glu Glu Ser Ser Ser Thr Val
165 170 175

Thr Asp Gln Asn Ser Ser Met Asp Asn Glu Asn His Leu Ile Asp Asn
180 185 190

Ile Tyr Asp Asp Asp Glu Leu Phe Ser Tyr Leu Trp Ser Asp Glu Thr
195 200 205

Thr Lys Asp Glu Ala Ser Trp Ser Asp Ser Asn Phe Gly Val Gly Gly
210 215 220

Thr Leu Tyr Asp His Asn Ile Ser Gly Ala Asp Ala Asp Phe Pro Ile
225 230 235 240

Trp Ser Pro Glu Arg Ile Asn Asp Glu Lys Met Phe Leu Asp Tyr Cys
245 250 255

MBI-20 Sequence Listing.ST25

```
Gln Asp Phe Gly Val His Asp Phe Gly Phe
            260                 265


<210>  5
<211>  2371
<212>  DNA
<213>  Arabidopsis thaliana

<220>
<221>  CDS
<222>  (172)..(2037)

<400>  5
gacattattt taagtgtgtt ctctctctgt cacactcaca aagctttata ctttctggct     60

actgcaagct catcagtgaa aagagcttaa accagagaga tctgataaga gaaattttag    120

agtctctctg cttcaacaag atctacatcg accaggagat tagaaagaat c atg ggt      177
                                                          Met Gly
                                                          1

tct aag cat aac cca cca ggg aat aac aga tcg aga agt aca cta tct      225
Ser Lys His Asn Pro Pro Gly Asn Asn Arg Ser Arg Ser Thr Leu Ser
        5                   10                  15

cta ctc gtt gtg gtt ggt tta tgt tgt ttc ttc tat ctt ctt gga gca      273
Leu Leu Val Val Val Gly Leu Cys Cys Phe Phe Tyr Leu Leu Gly Ala
    20                  25                  30

tgg caa aag agt ggg ttt ggt aaa gga gat agc ata gct atg gag att      321
Trp Gln Lys Ser Gly Phe Gly Lys Gly Asp Ser Ile Ala Met Glu Ile
35                  40                  45                  50

aca aag caa gcg cag tgt act gac att gtc act gat ctt gat ttt gaa      369
Thr Lys Gln Ala Gln Cys Thr Asp Ile Val Thr Asp Leu Asp Phe Glu
                55                  60                  65

cct cat cac aac aca gtg aag atc cca cat aaa gct gat ccc aaa cct      417
Pro His His Asn Thr Val Lys Ile Pro His Lys Ala Asp Pro Lys Pro
            70                  75                  80

gtt tct ttc aaa ccg tgt gat gtg aag ctc aag gat tac acg cct tgt      465
Val Ser Phe Lys Pro Cys Asp Val Lys Leu Lys Asp Tyr Thr Pro Cys
        85                  90                  95

caa gag caa gac cga gct atg aag ttc ccg aga gag aac atg att tac      513
Gln Glu Gln Asp Arg Ala Met Lys Phe Pro Arg Glu Asn Met Ile Tyr
    100                 105                 110

aga gag aga cat tgt cct cct gat aat gag aag ctg cgt tgt ctt gtt      561
Arg Glu Arg His Cys Pro Pro Asp Asn Glu Lys Leu Arg Cys Leu Val
115                 120                 125                 130

cca gct cct aaa ggg tat atg act cct ttc cct tgg cct aaa agc aga      609
Pro Ala Pro Lys Gly Tyr Met Thr Pro Phe Pro Trp Pro Lys Ser Arg
                135                 140                 145

gat tat gtt cac tat gct aat gct cct ttc aag agc ttg act gtc gaa      657
Asp Tyr Val His Tyr Ala Asn Ala Pro Phe Lys Ser Leu Thr Val Glu
            150                 155                 160

aaa gct gga cag aat tgg gtt cag ttt caa ggg aat gtg ttt aaa ttc      705
Lys Ala Gly Gln Asn Trp Val Gln Phe Gln Gly Asn Val Phe Lys Phe
        165                 170                 175

cct ggt gga gga act atg ttt cct caa ggt gct gat gcg tat att gaa      753
Pro Gly Gly Gly Thr Met Phe Pro Gln Gly Ala Asp Ala Tyr Ile Glu
    180                 185                 190

gag cta gct tct gtt atc cct atc aaa gat ggc tct gtt aga acc gca      801
Glu Leu Ala Ser Val Ile Pro Ile Lys Asp Gly Ser Val Arg Thr Ala
195                 200                 205                 210
```

```
ttg gac act gga tgt ggg gtt gct agt tgg ggt gct tat atg ctt aag      849
Leu Asp Thr Gly Cys Gly Val Ala Ser Trp Gly Ala Tyr Met Leu Lys
                215                 220                 225

agg aat gtt ttg act atg tcg ttt gcg cca agg gat aac cac gaa gca      897
Arg Asn Val Leu Thr Met Ser Phe Ala Pro Arg Asp Asn His Glu Ala
            230                 235                 240

caa gtc cag ttt gcg ctt gag aga ggt gtt cca gcg att atc gct gtt      945
Gln Val Gln Phe Ala Leu Glu Arg Gly Val Pro Ala Ile Ile Ala Val
        245                 250                 255

ctt gga tca atc ctt ctt cct tac cct gca aga gcc ttt gac atg gct      993
Leu Gly Ser Ile Leu Leu Pro Tyr Pro Ala Arg Ala Phe Asp Met Ala
    260                 265                 270

caa tgc tct cga tgc ttg ata cca tgg acc gca aac gag gga aca tac     1041
Gln Cys Ser Arg Cys Leu Ile Pro Trp Thr Ala Asn Glu Gly Thr Tyr
275                 280                 285                 290

tta atg gaa gta gat aga gtc ttg aga cct gga ggt tac tgg gtc tta     1089
Leu Met Glu Val Asp Arg Val Leu Arg Pro Gly Gly Tyr Trp Val Leu
                295                 300                 305

tcg ggt cct cca atc aac tgg aag aca tgg cac aag acg tgg aac cga     1137
Ser Gly Pro Pro Ile Asn Trp Lys Thr Trp His Lys Thr Trp Asn Arg
            310                 315                 320

act aaa gca gag cta aat gcc gag caa aag aga ata gag gga atc gca     1185
Thr Lys Ala Glu Leu Asn Ala Glu Gln Lys Arg Ile Glu Gly Ile Ala
        325                 330                 335

gag tcc tta tgc tgg gag aag aag tat gag aag gga gac att gca att     1233
Glu Ser Leu Cys Trp Glu Lys Lys Tyr Glu Lys Gly Asp Ile Ala Ile
    340                 345                 350

ttc aga aag aaa ata aac gat aga tca tgc gat aga tca aca ccg gtt     1281
Phe Arg Lys Lys Ile Asn Asp Arg Ser Cys Asp Arg Ser Thr Pro Val
355                 360                 365                 370

gac acc tgc aaa aga aag gac act gac gat gtc tgg tac aag gag ata     1329
Asp Thr Cys Lys Arg Lys Asp Thr Asp Asp Val Trp Tyr Lys Glu Ile
                375                 380                 385

gaa acg tgt gta aca cca ttc cct aaa gta tca aac gaa gaa gaa gtt     1377
Glu Thr Cys Val Thr Pro Phe Pro Lys Val Ser Asn Glu Glu Glu Val
            390                 395                 400

gct gga gga aag cta aag aag ttc ccc gag agg cta ttc gca gtg cct     1425
Ala Gly Gly Lys Leu Lys Lys Phe Pro Glu Arg Leu Phe Ala Val Pro
        405                 410                 415

cca agt atc tct aaa ggt ttg att aat ggc gtc gac gag gaa tca tac     1473
Pro Ser Ile Ser Lys Gly Leu Ile Asn Gly Val Asp Glu Glu Ser Tyr
    420                 425                 430

caa gaa gac atc aat cta tgg aag aag cga gtg acc gga tac aag aga     1521
Gln Glu Asp Ile Asn Leu Trp Lys Lys Arg Val Thr Gly Tyr Lys Arg
435                 440                 445                 450

att aac aga ctg ata ggt tcc acc aga tac cgt aat gtg atg gat atg     1569
Ile Asn Arg Leu Ile Gly Ser Thr Arg Tyr Arg Asn Val Met Asp Met
                455                 460                 465

aac gcc ggt ctt ggt gga ttc gct gct gcg ctt gaa tcg cct aaa tcg     1617
Asn Ala Gly Leu Gly Gly Phe Ala Ala Ala Leu Glu Ser Pro Lys Ser
            470                 475                 480

tgg gtt atg aat gtg att cca acc att aac aag aac aca ttg agt gtt     1665
Trp Val Met Asn Val Ile Pro Thr Ile Asn Lys Asn Thr Leu Ser Val
        485                 490                 495

gtt tat gag aga ggt ctc att ggt atc tat cat gac tgg tgt gaa ggc     1713
Val Tyr Glu Arg Gly Leu Ile Gly Ile Tyr His Asp Trp Cys Glu Gly
```

```
    500                 505                 510

ttt tca act tat cca aga aca tac gat ttc att cac gct agt ggt gtc     1761
Phe Ser Thr Tyr Pro Arg Thr Tyr Asp Phe Ile His Ala Ser Gly Val
515                 520                 525                 530

ttc agc ttg tat cag cac agc tgc aaa ctt gag gat att ctt ctt gaa     1809
Phe Ser Leu Tyr Gln His Ser Cys Lys Leu Glu Asp Ile Leu Leu Glu
                535                 540                 545

act gat cgg att tta cga ccg gaa ggg att gtg att ttc cgg gat gag     1857
Thr Asp Arg Ile Leu Arg Pro Glu Gly Ile Val Ile Phe Arg Asp Glu
            550                 555                 560

gtt gat gtt ttg aat gat gtg agg aag atc gtt gat gga atg aga tgg     1905
Val Asp Val Leu Asn Asp Val Arg Lys Ile Val Asp Gly Met Arg Trp
        565                 570                 575

gat act aag tta atg gat cat gaa gac ggt cct ctc gtg ccg gag aag     1953
Asp Thr Lys Leu Met Asp His Glu Asp Gly Pro Leu Val Pro Glu Lys
    580                 585                 590

att ctt gtc gcc acg aag cag tat tgg gta gcc ggc gac gat gga aac     2001
Ile Leu Val Ala Thr Lys Gln Tyr Trp Val Ala Gly Asp Asp Gly Asn
595                 600                 605                 610

aat tct ccg tcg tct tct aat agt gaa gaa gaa taa aacaaaaaca          2047
Asn Ser Pro Ser Ser Ser Asn Ser Glu Glu Glu
                615                 620

aaaaactcct caggttacta agcttgaagt gtagatctat tttacaacat ctggaaaatt   2107

cttatcaaaa aaggaaggaa tcagaatttc cattaaagaa aggtgtcaaa aaaaagttgt   2167

aaaactatat agtagtgatc aagacgaata tgtgcattta tgttttattt ttgttcccta   2227

gtttttaatt ttatttttt gaaggaagaa aaaattagtt ccatgtgttt ttgcaagata   2287

gttgaaacct tggacgcttg ttatgtatga tgcgatcttg acatttttta ataacagtta   2347

ttttaaataa atttatgata taaa                                         2371
```

<210> 6
<211> 621
<212> PRT
<213> Arabidopsis thaliana

<400> 6

```
Met Gly Ser Lys His Asn Pro Pro Gly Asn Asn Arg Ser Arg Ser Thr
1               5                   10                  15

Leu Ser Leu Leu Val Val Val Gly Leu Cys Cys Phe Phe Tyr Leu Leu
            20                  25                  30

Gly Ala Trp Gln Lys Ser Gly Phe Gly Lys Gly Asp Ser Ile Ala Met
        35                  40                  45

Glu Ile Thr Lys Gln Ala Gln Cys Thr Asp Ile Val Thr Asp Leu Asp
    50                  55                  60

Phe Glu Pro His His Asn Thr Val Lys Ile Pro His Lys Ala Asp Pro
65                  70                  75                  80

Lys Pro Val Ser Phe Lys Pro Cys Asp Val Lys Leu Lys Asp Tyr Thr
                85                  90                  95

Pro Cys Gln Glu Gln Asp Arg Ala Met Lys Phe Pro Arg Glu Asn Met
```

```
          100                 105                 110

Ile Tyr Arg Glu Arg His Cys Pro Pro Asp Asn Glu Lys Leu Arg Cys
        115                 120                 125

Leu Val Pro Ala Pro Lys Gly Tyr Met Thr Pro Phe Pro Trp Pro Lys
    130                 135                 140

Ser Arg Asp Tyr Val His Tyr Ala Asn Ala Pro Phe Lys Ser Leu Thr
145                 150                 155                 160

Val Glu Lys Ala Gly Gln Asn Trp Val Gln Phe Gln Gly Asn Val Phe
                165                 170                 175

Lys Phe Pro Gly Gly Gly Thr Met Phe Pro Gln Gly Ala Asp Ala Tyr
            180                 185                 190

Ile Glu Glu Leu Ala Ser Val Ile Pro Ile Lys Asp Gly Ser Val Arg
        195                 200                 205

Thr Ala Leu Asp Thr Gly Cys Gly Val Ala Ser Trp Gly Ala Tyr Met
    210                 215                 220

Leu Lys Arg Asn Val Leu Thr Met Ser Phe Ala Pro Arg Asp Asn His
225                 230                 235                 240

Glu Ala Gln Val Gln Phe Ala Leu Glu Arg Gly Val Pro Ala Ile Ile
                245                 250                 255

Ala Val Leu Gly Ser Ile Leu Leu Pro Tyr Pro Ala Arg Ala Phe Asp
            260                 265                 270

Met Ala Gln Cys Ser Arg Cys Leu Ile Pro Trp Thr Ala Asn Glu Gly
        275                 280                 285

Thr Tyr Leu Met Glu Val Asp Arg Val Leu Arg Pro Gly Gly Tyr Trp
    290                 295                 300

Val Leu Ser Gly Pro Pro Ile Asn Trp Lys Thr Trp His Lys Thr Trp
305                 310                 315                 320

Asn Arg Thr Lys Ala Glu Leu Asn Ala Glu Gln Lys Arg Ile Glu Gly
                325                 330                 335

Ile Ala Glu Ser Leu Cys Trp Glu Lys Lys Tyr Glu Lys Gly Asp Ile
            340                 345                 350

Ala Ile Phe Arg Lys Lys Ile Asn Asp Arg Ser Cys Asp Arg Ser Thr
        355                 360                 365

Pro Val Asp Thr Cys Lys Arg Lys Asp Thr Asp Asp Val Trp Tyr Lys
    370                 375                 380

Glu Ile Glu Thr Cys Val Thr Pro Phe Pro Lys Val Ser Asn Glu Glu
385                 390                 395                 400
```

Glu Val Ala Gly Gly Lys Leu Lys Lys Phe Pro Glu Arg Leu Phe Ala
405 410 415

Val Pro Pro Ser Ile Ser Lys Gly Leu Ile Asn Gly Val Asp Glu Glu
420 425 430

Ser Tyr Gln Glu Asp Ile Asn Leu Trp Lys Lys Arg Val Thr Gly Tyr
435 440 445

Lys Arg Ile Asn Arg Leu Ile Gly Ser Thr Arg Tyr Arg Asn Val Met
450 455 460

Asp Met Asn Ala Gly Leu Gly Gly Phe Ala Ala Ala Leu Glu Ser Pro
465 470 475 480

Lys Ser Trp Val Met Asn Val Ile Pro Thr Ile Asn Lys Asn Thr Leu
485 490 495

Ser Val Val Tyr Glu Arg Gly Leu Ile Gly Ile Tyr His Asp Trp Cys
500 505 510

Glu Gly Phe Ser Thr Tyr Pro Arg Thr Tyr Asp Phe Ile His Ala Ser
515 520 525

Gly Val Phe Ser Leu Tyr Gln His Ser Cys Lys Leu Glu Asp Ile Leu
530 535 540

Leu Glu Thr Asp Arg Ile Leu Arg Pro Glu Gly Ile Val Ile Phe Arg
545 550 555 560

Asp Glu Val Asp Val Leu Asn Asp Val Arg Lys Ile Val Asp Gly Met
565 570 575

Arg Trp Asp Thr Lys Leu Met Asp His Glu Asp Gly Pro Leu Val Pro
580 585 590

Glu Lys Ile Leu Val Ala Thr Lys Gln Tyr Trp Val Ala Gly Asp Asp
595 600 605

Gly Asn Asn Ser Pro Ser Ser Ser Asn Ser Glu Glu Glu
610 615 620

<210> 7
<211> 1764
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (1)..(1764)
<223> G307

<400> 7
atg aag aga gat cat cac caa ttc caa ggt cga ttg tcc aac cac ggg 48
Met Lys Arg Asp His His Gln Phe Gln Gly Arg Leu Ser Asn His Gly
1 5 10 15

act tct tct tct tca tca tca atc tct aaa gat aag atg atg atg gtg 96
Thr Ser Ser Ser Ser Ser Ser Ile Ser Lys Asp Lys Met Met Met Val
20 25 30

```
aaa aaa gaa gaa gac ggt gga ggt aac atg gac gac gag ctt ctc gct      144
Lys Lys Glu Glu Asp Gly Gly Gly Asn Met Asp Asp Glu Leu Leu Ala
        35                  40                  45

gtt tta ggt tac aaa gtt agg tca tcg gag atg gcg gag gtt gct ttg      192
Val Leu Gly Tyr Lys Val Arg Ser Ser Glu Met Ala Glu Val Ala Leu
    50                  55                  60

aaa ctc gaa caa tta gag acg atg atg agt aat gtt caa gaa gat ggt      240
Lys Leu Glu Gln Leu Glu Thr Met Met Ser Asn Val Gln Glu Asp Gly
65                  70                  75                  80

tta tct cat ctc gcg acg gat act gtt cat tat aat ccg tcg gag ctt      288
Leu Ser His Leu Ala Thr Asp Thr Val His Tyr Asn Pro Ser Glu Leu
                85                  90                  95

tat tct tgg ctt gat aat atg ctc tct gag ctt aat cct cct cct ctt      336
Tyr Ser Trp Leu Asp Asn Met Leu Ser Glu Leu Asn Pro Pro Pro Leu
            100                 105                 110

ccg gcg agt tct aac ggt tta gat ccg gtt ctt cct tcg ccg gag att      384
Pro Ala Ser Ser Asn Gly Leu Asp Pro Val Leu Pro Ser Pro Glu Ile
        115                 120                 125

tgt ggt ttt ccg gct tcg gat tat gac ctt aaa gtc att ccc gga aac      432
Cys Gly Phe Pro Ala Ser Asp Tyr Asp Leu Lys Val Ile Pro Gly Asn
    130                 135                 140

gcg att tat cag ttt ccg gcg att gat tct tcg tct tcg tcg aat aat      480
Ala Ile Tyr Gln Phe Pro Ala Ile Asp Ser Ser Ser Ser Ser Asn Asn
145                 150                 155                 160

cag aac aag cgt ttg aaa tca tgc tcg agt cct gat tct atg gtt aca      528
Gln Asn Lys Arg Leu Lys Ser Cys Ser Ser Pro Asp Ser Met Val Thr
                165                 170                 175

tcg act tcg acg ggt acg cag att ggt gga gtc ata gga acg acg gtg      576
Ser Thr Ser Thr Gly Thr Gln Ile Gly Gly Val Ile Gly Thr Thr Val
            180                 185                 190

acg aca acc acc acg aca acg acg gcg gcg gct gag tca act cgt tct      624
Thr Thr Thr Thr Thr Thr Thr Ala Ala Ala Glu Ser Thr Arg Ser
        195                 200                 205

gtt atc ctg gtt gac tcg caa gag aac ggt gtt cgt tta gtc cac gcg      672
Val Ile Leu Val Asp Ser Gln Glu Asn Gly Val Arg Leu Val His Ala
    210                 215                 220

ctt atg gct tgt gca gaa gca atc cag cag aac aat ttg act cta gcg      720
Leu Met Ala Cys Ala Glu Ala Ile Gln Gln Asn Asn Leu Thr Leu Ala
225                 230                 235                 240

gaa gct ctt gtg aag caa atc gga tgc tta gct gtg tct caa gcc gga      768
Glu Ala Leu Val Lys Gln Ile Gly Cys Leu Ala Val Ser Gln Ala Gly
                245                 250                 255

gct atg aga aaa gtg gct act tac ttc gcc gaa gct tta gct cgg cgg      816
Ala Met Arg Lys Val Ala Thr Tyr Phe Ala Glu Ala Leu Ala Arg Arg
            260                 265                 270

atc tac cgt ctc tct ccg ccg cag aat cag atc gat cat tgt ctc tcc      864
Ile Tyr Arg Leu Ser Pro Pro Gln Asn Gln Ile Asp His Cys Leu Ser
        275                 280                 285

gat act ctt cag atg cac ttt tac gag act tgt cct tat ctt aaa ttc      912
Asp Thr Leu Gln Met His Phe Tyr Glu Thr Cys Pro Tyr Leu Lys Phe
    290                 295                 300

gct cac ttc acg gcg aac caa gcg att ctc gaa gct ttt gaa ggt aag      960
Ala His Phe Thr Ala Asn Gln Ala Ile Leu Glu Ala Phe Glu Gly Lys
305                 310                 315                 320

aag aga gta cac gtc att gat ttc tcg atg aac caa ggt ctt caa tgg     1008
Lys Arg Val His Val Ile Asp Phe Ser Met Asn Gln Gly Leu Gln Trp
```

```
                325                 330                 335

cct gcg ctt atg caa gct ctt gcg ctt cga gaa gga ggt cct cca act     1056
Pro Ala Leu Met Gln Ala Leu Ala Leu Arg Glu Gly Gly Pro Pro Thr
            340                 345                 350

ttc cgg tta acc gga att ggt cca ccg gcg ccg gat aat tct gat cat     1104
Phe Arg Leu Thr Gly Ile Gly Pro Pro Ala Pro Asp Asn Ser Asp His
        355                 360                 365

ctt cat gaa gtt ggt tgt aaa tta gct cag ctt gcg gag gcg att cac     1152
Leu His Glu Val Gly Cys Lys Leu Ala Gln Leu Ala Glu Ala Ile His
    370                 375                 380

gta gaa ttc gaa tac cgt gga ttc gtt gct aac agc tta gcc gat ctc     1200
Val Glu Phe Glu Tyr Arg Gly Phe Val Ala Asn Ser Leu Ala Asp Leu
385                 390                 395                 400

gat gct tcg atg ctt gag ctt aga ccg agc gat acg gaa gct gtt gcg     1248
Asp Ala Ser Met Leu Glu Leu Arg Pro Ser Asp Thr Glu Ala Val Ala
                405                 410                 415

gtg aac tct gtt ttt gag cta cat aag ctc tta ggt cgt ccc ggt ggg     1296
Val Asn Ser Val Phe Glu Leu His Lys Leu Leu Gly Arg Pro Gly Gly
            420                 425                 430

ata gag aaa gtt ctc ggc gtt gtg aaa cag att aaa ccg gtg att ttc     1344
Ile Glu Lys Val Leu Gly Val Val Lys Gln Ile Lys Pro Val Ile Phe
        435                 440                 445

acg gtg gtt gag caa gaa tcg aac cat aac gga ccg gtt ttc tta gac     1392
Thr Val Val Glu Gln Glu Ser Asn His Asn Gly Pro Val Phe Leu Asp
    450                 455                 460

cgg ttt act gaa tcg tta cat tat tat tcg act ctg ttt gat tcg ttg     1440
Arg Phe Thr Glu Ser Leu His Tyr Tyr Ser Thr Leu Phe Asp Ser Leu
465                 470                 475                 480

gaa gga gtt ccg aat agt caa gac aaa gtc atg tct gaa gtt tac tta     1488
Glu Gly Val Pro Asn Ser Gln Asp Lys Val Met Ser Glu Val Tyr Leu
                485                 490                 495

ggg aaa cag att tgt aat ctg gtg gct tgt gaa ggt cct gac aga gtc     1536
Gly Lys Gln Ile Cys Asn Leu Val Ala Cys Glu Gly Pro Asp Arg Val
            500                 505                 510

gag aga cac gaa acg ttg agt caa tgg gga aac cgg ttt ggt tcg tcc     1584
Glu Arg His Glu Thr Leu Ser Gln Trp Gly Asn Arg Phe Gly Ser Ser
        515                 520                 525

ggt tta gcg ccg gca cat ctt ggg tct aac gcg ttt aag caa gcg agt     1632
Gly Leu Ala Pro Ala His Leu Gly Ser Asn Ala Phe Lys Gln Ala Ser
    530                 535                 540

atg ctt ttg tct gtg ttt aat agt ggc caa ggt tat cgt gtg gag gag     1680
Met Leu Leu Ser Val Phe Asn Ser Gly Gln Gly Tyr Arg Val Glu Glu
545                 550                 555                 560

agt aat gga tgt ttg atg ttg ggt tgg cac act cgc cca ctc att acc     1728
Ser Asn Gly Cys Leu Met Leu Gly Trp His Thr Arg Pro Leu Ile Thr
                565                 570                 575

acc tcc gct tgg aaa ctc tcg acg gcg gcg cac tga                     1764
Thr Ser Ala Trp Lys Leu Ser Thr Ala Ala His
            580                 585
```

<210> 8
<211> 587
<212> PRT
<213> Arabidopsis thaliana

<400> 8

```
Met Lys Arg Asp His His Gln Phe Gln Gly Arg Leu Ser Asn His Gly
```

MBI-20 Sequence Listing.ST25

```
1               5                  10                  15

Thr Ser Ser Ser Ser Ser Ser Ile Ser Lys Asp Lys Met Met Met Val
            20                  25                  30

Lys Lys Glu Glu Asp Gly Gly Gly Asn Met Asp Asp Glu Leu Leu Ala
        35                  40                  45

Val Leu Gly Tyr Lys Val Arg Ser Ser Glu Met Ala Glu Val Ala Leu
    50                  55                  60

Lys Leu Glu Gln Leu Glu Thr Met Met Ser Asn Val Gln Glu Asp Gly
65                  70                  75                  80

Leu Ser His Leu Ala Thr Asp Thr Val His Tyr Asn Pro Ser Glu Leu
                85                  90                  95

Tyr Ser Trp Leu Asp Asn Met Leu Ser Glu Leu Asn Pro Pro Pro Leu
            100                 105                 110

Pro Ala Ser Ser Asn Gly Leu Asp Pro Val Leu Pro Ser Pro Glu Ile
        115                 120                 125

Cys Gly Phe Pro Ala Ser Asp Tyr Asp Leu Lys Val Ile Pro Gly Asn
    130                 135                 140

Ala Ile Tyr Gln Phe Pro Ala Ile Asp Ser Ser Ser Ser Ser Asn Asn
145                 150                 155                 160

Gln Asn Lys Arg Leu Lys Ser Cys Ser Ser Pro Asp Ser Met Val Thr
                165                 170                 175

Ser Thr Ser Thr Gly Thr Gln Ile Gly Gly Val Ile Gly Thr Thr Val
            180                 185                 190

Thr Thr Thr Thr Thr Thr Thr Thr Ala Ala Ala Glu Ser Thr Arg Ser
        195                 200                 205

Val Ile Leu Val Asp Ser Gln Glu Asn Gly Val Arg Leu Val His Ala
    210                 215                 220

Leu Met Ala Cys Ala Glu Ala Ile Gln Gln Asn Asn Leu Thr Leu Ala
225                 230                 235                 240

Glu Ala Leu Val Lys Gln Ile Gly Cys Leu Ala Val Ser Gln Ala Gly
                245                 250                 255

Ala Met Arg Lys Val Ala Thr Tyr Phe Ala Glu Ala Leu Ala Arg Arg
            260                 265                 270

Ile Tyr Arg Leu Ser Pro Pro Gln Asn Gln Ile Asp His Cys Leu Ser
        275                 280                 285

Asp Thr Leu Gln Met His Phe Tyr Glu Thr Cys Pro Tyr Leu Lys Phe
    290                 295                 300
```

```
Ala His Phe Thr Ala Asn Gln Ala Ile Leu Glu Ala Phe Glu Gly Lys
305                 310                 315                 320

Lys Arg Val His Val Ile Asp Phe Ser Met Asn Gln Gly Leu Gln Trp
                325                 330                 335

Pro Ala Leu Met Gln Ala Leu Ala Leu Arg Glu Gly Gly Pro Pro Thr
            340                 345                 350

Phe Arg Leu Thr Gly Ile Gly Pro Pro Ala Pro Asp Asn Ser Asp His
        355                 360                 365

Leu His Glu Val Gly Cys Lys Leu Ala Gln Leu Ala Glu Ala Ile His
    370                 375                 380

Val Glu Phe Glu Tyr Arg Gly Phe Val Ala Asn Ser Leu Ala Asp Leu
385                 390                 395                 400

Asp Ala Ser Met Leu Glu Leu Arg Pro Ser Asp Thr Glu Ala Val Ala
                405                 410                 415

Val Asn Ser Val Phe Glu Leu His Lys Leu Leu Gly Arg Pro Gly Gly
            420                 425                 430

Ile Glu Lys Val Leu Gly Val Val Lys Gln Ile Lys Pro Val Ile Phe
        435                 440                 445

Thr Val Val Glu Gln Glu Ser Asn His Asn Gly Pro Val Phe Leu Asp
    450                 455                 460

Arg Phe Thr Glu Ser Leu His Tyr Tyr Ser Thr Leu Phe Asp Ser Leu
465                 470                 475                 480

Glu Gly Val Pro Asn Ser Gln Asp Lys Val Met Ser Glu Val Tyr Leu
                485                 490                 495

Gly Lys Gln Ile Cys Asn Leu Val Ala Cys Glu Gly Pro Asp Arg Val
            500                 505                 510

Glu Arg His Glu Thr Leu Ser Gln Trp Gly Asn Arg Phe Gly Ser Ser
        515                 520                 525

Gly Leu Ala Pro Ala His Leu Gly Ser Asn Ala Phe Lys Gln Ala Ser
    530                 535                 540

Met Leu Leu Ser Val Phe Asn Ser Gly Gln Gly Tyr Arg Val Glu Glu
545                 550                 555                 560

Ser Asn Gly Cys Leu Met Leu Gly Trp His Thr Arg Pro Leu Ile Thr
                565                 570                 575

Thr Ser Ala Trp Lys Leu Ser Thr Ala Ala His
            580                 585
```

<210> 9
<211> 825
<212> DNA

MBI-20 Sequence Listing.ST25

<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (1)..(825)
<223> G346

<400> 9

```
atg gaa atg gaa tca ttc atg gac gac ctt ttg aac ttc tct gta ccg       48
Met Glu Met Glu Ser Phe Met Asp Asp Leu Leu Asn Phe Ser Val Pro
1               5                  10                  15

gaa gag gaa gaa gac gac gac gaa cat acg caa cca ccg agg aat att       96
Glu Glu Glu Glu Asp Asp Asp Glu His Thr Gln Pro Pro Arg Asn Ile
            20                  25                  30

act cgc cgg aaa act gga tta cgg cca aca gac tcc ttc ggt ctc ttt      144
Thr Arg Arg Lys Thr Gly Leu Arg Pro Thr Asp Ser Phe Gly Leu Phe
        35                  40                  45

aat acc gac gac ctt gga gtg gtt gaa gaa gag gat ttg gaa tgg att      192
Asn Thr Asp Asp Leu Gly Val Val Glu Glu Glu Asp Leu Glu Trp Ile
    50                  55                  60

tca aac aaa aat gct ttt ccg gtg att gaa aca ttc gtc ggt gta tta      240
Ser Asn Lys Asn Ala Phe Pro Val Ile Glu Thr Phe Val Gly Val Leu
65                  70                  75                  80

ccg tcg gag cat ttt cct ata acg tct ctt ctg gaa aga gaa gcg act      288
Pro Ser Glu His Phe Pro Ile Thr Ser Leu Leu Glu Arg Glu Ala Thr
                85                  90                  95

gag gta aaa cag ctg agt ccg gtt tca gta ctt gag acg agt agc cat      336
Glu Val Lys Gln Leu Ser Pro Val Ser Val Leu Glu Thr Ser Ser His
            100                 105                 110

agc tcc aca acg act acc tca aac agt agc ggc gga agt aac gga agc      384
Ser Ser Thr Thr Thr Thr Ser Asn Ser Ser Gly Gly Ser Asn Gly Ser
        115                 120                 125

acg gcc gtg gct acg acc acc acc act cca aca ata atg agc tgt tgc      432
Thr Ala Val Ala Thr Thr Thr Thr Thr Pro Thr Ile Met Ser Cys Cys
    130                 135                 140

gtt ggt ttt aaa gcg ccg gct aaa gcg aga agc aag cgt cgt cgt aca      480
Val Gly Phe Lys Ala Pro Ala Lys Ala Arg Ser Lys Arg Arg Arg Thr
145                 150                 155                 160

gga cgc cgt gat tta cga gtt ttg tgg aca gga aac gag caa gga gga      528
Gly Arg Arg Asp Leu Arg Val Leu Trp Thr Gly Asn Glu Gln Gly Gly
                165                 170                 175

ata cag aag aag aag acg atg act gtg gcg gcg gct gcg ttg att atg      576
Ile Gln Lys Lys Lys Thr Met Thr Val Ala Ala Ala Ala Leu Ile Met
            180                 185                 190

gga agg aag tgt caa cac tgt gga gcg gag aag act ccg caa tgg agg      624
Gly Arg Lys Cys Gln His Cys Gly Ala Glu Lys Thr Pro Gln Trp Arg
        195                 200                 205

gca gga cca gcg ggg cct aag act ctg tgt aac gct tgt ggc gtg agg      672
Ala Gly Pro Ala Gly Pro Lys Thr Leu Cys Asn Ala Cys Gly Val Arg
    210                 215                 220

tat aag tcc ggg agg cta gtt ccg gag tat cgt cca gcg aac agt cca      720
Tyr Lys Ser Gly Arg Leu Val Pro Glu Tyr Arg Pro Ala Asn Ser Pro
225                 230                 235                 240

act ttc acg gcg gag tta cat tcg aat tct cac cgg aag att gta gag      768
Thr Phe Thr Ala Glu Leu His Ser Asn Ser His Arg Lys Ile Val Glu
                245                 250                 255

atg agg aag cag tat cag tcc ggt gac ggt gac ggt gat cgg aaa gat      816
Met Arg Lys Gln Tyr Gln Ser Gly Asp Gly Asp Gly Asp Arg Lys Asp
```

260 265 270

tgt gga taa 825
Cys Gly

<210> 10
<211> 274
<212> PRT
<213> Arabidopsis thaliana

<400> 10

Met Glu Met Glu Ser Phe Met Asp Asp Leu Leu Asn Phe Ser Val Pro
1 5 10 15

Glu Glu Glu Glu Asp Asp Asp Glu His Thr Gln Pro Pro Arg Asn Ile
20 25 30

Thr Arg Arg Lys Thr Gly Leu Arg Pro Thr Asp Ser Phe Gly Leu Phe
35 40 45

Asn Thr Asp Asp Leu Gly Val Val Glu Glu Glu Asp Leu Glu Trp Ile
50 55 60

Ser Asn Lys Asn Ala Phe Pro Val Ile Glu Thr Phe Val Gly Val Leu
65 70 75 80

Pro Ser Glu His Phe Pro Ile Thr Ser Leu Leu Glu Arg Glu Ala Thr
85 90 95

Glu Val Lys Gln Leu Ser Pro Val Ser Val Leu Glu Thr Ser Ser His
100 105 110

Ser Ser Thr Thr Thr Thr Ser Asn Ser Ser Gly Gly Ser Asn Gly Ser
115 120 125

Thr Ala Val Ala Thr Thr Thr Thr Thr Pro Thr Ile Met Ser Cys Cys
130 135 140

Val Gly Phe Lys Ala Pro Ala Lys Ala Arg Ser Lys Arg Arg Arg Thr
145 150 155 160

Gly Arg Arg Asp Leu Arg Val Leu Trp Thr Gly Asn Glu Gln Gly Gly
165 170 175

Ile Gln Lys Lys Lys Thr Met Thr Val Ala Ala Ala Ala Leu Ile Met
180 185 190

Gly Arg Lys Cys Gln His Cys Gly Ala Glu Lys Thr Pro Gln Trp Arg
195 200 205

Ala Gly Pro Ala Gly Pro Lys Thr Leu Cys Asn Ala Cys Gly Val Arg
210 215 220

Tyr Lys Ser Gly Arg Leu Val Pro Glu Tyr Arg Pro Ala Asn Ser Pro
225 230 235 240

Thr Phe Thr Ala Glu Leu His Ser Asn Ser His Arg Lys Ile Val Glu

245 250 255

Met Arg Lys Gln Tyr Gln Ser Gly Asp Gly Asp Gly Asp Arg Lys Asp
260 265 270

Cys Gly

<210> 11
<211> 1226
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (248)..(1039)
<223> G598

<400> 11
gtccgttgtc atattttaaa tttatcacct tcttgagaat tccacatttt tatccttttt 60

gtcatgtagt gtatattttt tcctctaacc taattaaaat caaaacaaaa tcctttgacc 120

caattagctt cgcgatatat cagaagagat caaactactt tgatcagacc atgatcttct 180

tcttcttctt cttcttcttc ttcttctttt tagacgatca caattcctaa accctatttc 240

tcagatt atg ctg act ctt tac cat caa gaa agg tca ccg gac gcc aca 289
Met Leu Thr Leu Tyr His Gln Glu Arg Ser Pro Asp Ala Thr
1 5 10

agt aat gat cgc gat gag acg cca gag act gtg gtt aga gaa gtc cac 337
Ser Asn Asp Arg Asp Glu Thr Pro Glu Thr Val Val Arg Glu Val His
15 20 25 30

gcg cta act cca gcg ccg gag gat aat tcc cgg acg atg acg gcg acg 385
Ala Leu Thr Pro Ala Pro Glu Asp Asn Ser Arg Thr Met Thr Ala Thr
35 40 45

cta cct cca ccg cct gct ttc cga ggc tat ttt tct cct cca agg tca 433
Leu Pro Pro Pro Pro Ala Phe Arg Gly Tyr Phe Ser Pro Pro Arg Ser
50 55 60

gcg acg acg atg agc gaa gga gag aac ttc aca act ata agc aga gag 481
Ala Thr Thr Met Ser Glu Gly Glu Asn Phe Thr Thr Ile Ser Arg Glu
65 70 75

ttc aac gct cta gtc atc gcc gga tcc tcc atg gag aac aac gaa cta 529
Phe Asn Ala Leu Val Ile Ala Gly Ser Ser Met Glu Asn Asn Glu Leu
80 85 90

atg act cgt gac gtc acg cag cgt gaa gat gag aga caa gac gag ttg 577
Met Thr Arg Asp Val Thr Gln Arg Glu Asp Glu Arg Gln Asp Glu Leu
95 100 105 110

atg aga atc cac gag gac acg gat cat gaa gag gaa acg aat cct tta 625
Met Arg Ile His Glu Asp Thr Asp His Glu Glu Glu Thr Asn Pro Leu
115 120 125

gca atc gtg ccg gat cag tat cct ggt tcg ggt ttg gat cct gga agt 673
Ala Ile Val Pro Asp Gln Tyr Pro Gly Ser Gly Leu Asp Pro Gly Ser
130 135 140

gat aat ggg ccg ggt cag agt cgg gtt ggg tcg acg gtg caa aga gtt 721
Asp Asn Gly Pro Gly Gln Ser Arg Val Gly Ser Thr Val Gln Arg Val
145 150 155

aag agg gaa gag gtg gaa gcg aag ata acg gcg tgg cag acg gca aaa 769
Lys Arg Glu Glu Val Glu Ala Lys Ile Thr Ala Trp Gln Thr Ala Lys
160 165 170

```
ctg gct aag att aat aac agg ttt aag agg gaa gac gcc gtt att aac      817
Leu Ala Lys Ile Asn Asn Arg Phe Lys Arg Glu Asp Ala Val Ile Asn
175                 180                 185                 190

ggt tgg ttt aat gaa caa gtt aac aag gcc aac tct tgg atg aag aaa      865
Gly Trp Phe Asn Glu Gln Val Asn Lys Ala Asn Ser Trp Met Lys Lys
                195                 200                 205

att gag tat aat gta ggt tca ttc aac aat cgt cta aat gag gaa gct      913
Ile Glu Tyr Asn Val Gly Ser Phe Asn Asn Arg Leu Asn Glu Glu Ala
            210                 215                 220

aga gga gag aaa agc aaa agc gat gga gaa aac gca aaa caa tgt ggc      961
Arg Gly Glu Lys Ser Lys Ser Asp Gly Glu Asn Ala Lys Gln Cys Gly
        225                 230                 235

gaa agc gca gag gaa agc gga gga gag aag agc gac ggc aga ggc aaa     1009
Glu Ser Ala Glu Glu Ser Gly Gly Glu Lys Ser Asp Gly Arg Gly Lys
    240                 245                 250

gag agg gac aga ggt tgc aaa agt agt tga agttgctaat ctcatgagag       1059
Glu Arg Asp Arg Gly Cys Lys Ser Ser
255                 260

cccttggacg tcctcctgcc aaacgctcct tcttctcttt ctcctaattt ttagttatat   1119

caaaccatta aattaaacag tactcgttat atatctagtt agtaaacaaa ggggcagttt   1179

tatagctcat gtacacataa ttgagagtgt agtactgttg tgtcaaa                 1226
```

<210> 12
<211> 263
<212> PRT
<213> Arabidopsis thaliana

<400> 12

```
Met Leu Thr Leu Tyr His Gln Glu Arg Ser Pro Asp Ala Thr Ser Asn
1               5                   10                  15

Asp Arg Asp Glu Thr Pro Glu Thr Val Val Arg Glu Val His Ala Leu
            20                  25                  30

Thr Pro Ala Pro Glu Asp Asn Ser Arg Thr Met Thr Ala Thr Leu Pro
        35                  40                  45

Pro Pro Pro Ala Phe Arg Gly Tyr Phe Ser Pro Pro Arg Ser Ala Thr
    50                  55                  60

Thr Met Ser Glu Gly Glu Asn Phe Thr Thr Ile Ser Arg Glu Phe Asn
65                  70                  75                  80

Ala Leu Val Ile Ala Gly Ser Ser Met Glu Asn Asn Glu Leu Met Thr
                85                  90                  95

Arg Asp Val Thr Gln Arg Glu Asp Glu Arg Gln Asp Glu Leu Met Arg
            100                 105                 110

Ile His Glu Asp Thr Asp His Glu Glu Glu Thr Asn Pro Leu Ala Ile
        115                 120                 125

Val Pro Asp Gln Tyr Pro Gly Ser Gly Leu Asp Pro Gly Ser Asp Asn
    130                 135                 140

Gly Pro Gly Gln Ser Arg Val Gly Ser Thr Val Gln Arg Val Lys Arg
```

```
145                 150                 155                 160

Glu Glu Val Glu Ala Lys Ile Thr Ala Trp Gln Thr Ala Lys Leu Ala
                165                 170                 175

Lys Ile Asn Asn Arg Phe Lys Arg Glu Asp Ala Val Ile Asn Gly Trp
            180                 185                 190

Phe Asn Glu Gln Val Asn Lys Ala Asn Ser Trp Met Lys Lys Ile Glu
        195                 200                 205

Tyr Asn Val Gly Ser Phe Asn Asn Arg Leu Asn Glu Glu Ala Arg Gly
    210                 215                 220

Glu Lys Ser Lys Ser Asp Gly Glu Asn Ala Lys Gln Cys Gly Glu Ser
225                 230                 235                 240

Ala Glu Glu Ser Gly Gly Glu Lys Ser Asp Gly Arg Gly Lys Glu Arg
                245                 250                 255

Asp Arg Gly Cys Lys Ser Ser
            260
```

<210> 13
<211> 1263
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (72)..(1076)
<223> G605

<400> 13

```
aattccatcc taataatttt caaagcttta attctaagaa ataatatcta caagaaaata     60

ttatctcatg t atg gag act acc gga gaa gtt gtt aaa aca acc acc ggg     110
             Met Glu Thr Thr Gly Glu Val Val Lys Thr Thr Thr Gly
             1               5                   10

agc gac gga ggc gtt acg gtg gtg aga tcc aac gcg ccg tca gac ttc     158
Ser Asp Gly Gly Val Thr Val Val Arg Ser Asn Ala Pro Ser Asp Phe
    15                  20                  25

cac atg gct ccg agg tca gaa act tca aac aca cct ccc aac tcc gtc     206
His Met Ala Pro Arg Ser Glu Thr Ser Asn Thr Pro Pro Asn Ser Val
30                  35                  40                  45

gct cct cct cct cct cca ccg ccg caa aac tcc ttt act ccg tcg gcg     254
Ala Pro Pro Pro Pro Pro Pro Pro Gln Asn Ser Phe Thr Pro Ser Ala
                50                  55                  60

gct atg gat ggt ttc tca agc gga ccg ata aag aag aga cgt ggg cgc     302
Ala Met Asp Gly Phe Ser Ser Gly Pro Ile Lys Lys Arg Arg Gly Arg
            65                  70                  75

cct agg aag tac gga cac gac gga gca gcg gtg acg cta tct ccg aat     350
Pro Arg Lys Tyr Gly His Asp Gly Ala Ala Val Thr Leu Ser Pro Asn
        80                  85                  90

ccg ata tca tca gcc gca cca acg act tct cac gtc atc gat ttc tcg     398
Pro Ile Ser Ser Ala Ala Pro Thr Thr Ser His Val Ile Asp Phe Ser
    95                  100                 105

acg aca tcg gag aaa cgt ggc aaa atg aaa cca gca act cca act cca     446
Thr Thr Ser Glu Lys Arg Gly Lys Met Lys Pro Ala Thr Pro Thr Pro
```

```
110                 115                 120                 125

agc tca ttc atc agg cca aag tac cag gtc gag aat tta ggt gaa tgg      494
Ser Ser Phe Ile Arg Pro Lys Tyr Gln Val Glu Asn Leu Gly Glu Trp
                130                 135                 140

tct cct tcc tct gcc gcc gct aat ttc acg ccg cat att att acg gtg      542
Ser Pro Ser Ser Ala Ala Ala Asn Phe Thr Pro His Ile Ile Thr Val
            145                 150                 155

aat gca ggc gag gac gtt acg aag agg ata ata tca ttt tct caa caa      590
Asn Ala Gly Glu Asp Val Thr Lys Arg Ile Ile Ser Phe Ser Gln Gln
        160                 165                 170

ggg tct cta gct att tgc gtt tta tgc gca aac ggt gtc gtt tcg agc      638
Gly Ser Leu Ala Ile Cys Val Leu Cys Ala Asn Gly Val Val Ser Ser
    175                 180                 185

gtt aca ctt cgt cag cct gat tca tct ggt ggt aca ttg acc tat gag      686
Val Thr Leu Arg Gln Pro Asp Ser Ser Gly Gly Thr Leu Thr Tyr Glu
190                 195                 200                 205

ggt cgg ttt gag ata ttg tca cta tct gga aca ttc atg cct agt gac      734
Gly Arg Phe Glu Ile Leu Ser Leu Ser Gly Thr Phe Met Pro Ser Asp
                210                 215                 220

tca gac ggg aca cga agc aga aca ggc ggg atg agc gtg tcg ctt gct      782
Ser Asp Gly Thr Arg Ser Arg Thr Gly Gly Met Ser Val Ser Leu Ala
            225                 230                 235

agc cct gat gga cgt gta gta ggt ggt ggt gtt gct ggc ttg ctg gtt      830
Ser Pro Asp Gly Arg Val Val Gly Gly Gly Val Ala Gly Leu Leu Val
        240                 245                 250

gca gcc act cct att caa gtg gtt gta gga act ttc tta ggt gga aca      878
Ala Ala Thr Pro Ile Gln Val Val Val Gly Thr Phe Leu Gly Gly Thr
    255                 260                 265

aac cag caa gaa cag aca ccg aag ccg cat aac cac aac ttc atg tct      926
Asn Gln Gln Glu Gln Thr Pro Lys Pro His Asn His Asn Phe Met Ser
270                 275                 280                 285

tct cca tta atg cca act tct tcg aat gta gct gat cat cga acc atc      974
Ser Pro Leu Met Pro Thr Ser Ser Asn Val Ala Asp His Arg Thr Ile
                290                 295                 300

cgt ccc atg aca tct agt ctc ccg atc agt aca tgg aca ccg tct ttt     1022
Arg Pro Met Thr Ser Ser Leu Pro Ile Ser Thr Trp Thr Pro Ser Phe
            305                 310                 315

cct tct gat tca cga cac aag cat tct cat gac ttt aat atc act ttg     1070
Pro Ser Asp Ser Arg His Lys His Ser His Asp Phe Asn Ile Thr Leu
        320                 325                 330

acg tga tttcttcctt gaagaactcg tagatcctct gtattttggt ttccagttta     1126
Thr

gggctctaca tgttagactc tcaaagtcta ggtgttatgt tggtctgtca cttaggattg  1186

tcacttagga ttgttagacc atctccatca atggtttctc attgagaaac tgttcaatat  1246

aaaaataaaa tataatc                                                 1263
```

<210> 14
<211> 334
<212> PRT
<213> Arabidopsis thaliana

<400> 14

```
Met Glu Thr Thr Gly Glu Val Val Lys Thr Thr Thr Gly Ser Asp Gly
1               5                   10                  15
```

MBI-20 Sequence Listing.ST25

```
Gly Val Thr Val Val Arg Ser Asn Ala Pro Ser Asp Phe His Met Ala
           20                  25                  30

Pro Arg Ser Glu Thr Ser Asn Thr Pro Pro Asn Ser Val Ala Pro Pro
       35                  40                  45

Pro Pro Pro Pro Pro Gln Asn Ser Phe Thr Pro Ser Ala Ala Met Asp
   50                  55                  60

Gly Phe Ser Ser Gly Pro Ile Lys Lys Arg Arg Gly Arg Pro Arg Lys
65                  70                  75                  80

Tyr Gly His Asp Gly Ala Ala Val Thr Leu Ser Pro Asn Pro Ile Ser
               85                  90                  95

Ser Ala Ala Pro Thr Thr Ser His Val Ile Asp Phe Ser Thr Thr Ser
           100                 105                 110

Glu Lys Arg Gly Lys Met Lys Pro Ala Thr Pro Thr Pro Ser Ser Phe
       115                 120                 125

Ile Arg Pro Lys Tyr Gln Val Glu Asn Leu Gly Glu Trp Ser Pro Ser
   130                 135                 140

Ser Ala Ala Ala Asn Phe Thr Pro His Ile Ile Thr Val Asn Ala Gly
145                 150                 155                 160

Glu Asp Val Thr Lys Arg Ile Ile Ser Phe Ser Gln Gln Gly Ser Leu
               165                 170                 175

Ala Ile Cys Val Leu Cys Ala Asn Gly Val Val Ser Ser Val Thr Leu
           180                 185                 190

Arg Gln Pro Asp Ser Ser Gly Gly Thr Leu Thr Tyr Glu Gly Arg Phe
       195                 200                 205

Glu Ile Leu Ser Leu Ser Gly Thr Phe Met Pro Ser Asp Ser Asp Gly
   210                 215                 220

Thr Arg Ser Arg Thr Gly Gly Met Ser Val Ser Leu Ala Ser Pro Asp
225                 230                 235                 240

Gly Arg Val Val Gly Gly Gly Val Ala Gly Leu Leu Val Ala Ala Thr
               245                 250                 255

Pro Ile Gln Val Val Val Gly Thr Phe Leu Gly Gly Thr Asn Gln Gln
           260                 265                 270

Glu Gln Thr Pro Lys Pro His Asn His Asn Phe Met Ser Ser Pro Leu
       275                 280                 285

Met Pro Thr Ser Ser Asn Val Ala Asp His Arg Thr Ile Arg Pro Met
   290                 295                 300

Thr Ser Ser Leu Pro Ile Ser Thr Trp Thr Pro Ser Phe Pro Ser Asp
305                 310                 315                 320
```

MBI-20 Sequence Listing.ST25

```
Ser Arg His Lys His Ser His Asp Phe Asn Ile Thr Leu Thr
                325                 330
```

<210> 15
<211> 1057
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (54)..(914)
<223> G777

<400> 15

```
gtggctctct ctttatcttt cttggagttt agttagagat tttaacgttg caa atg        56
                                                          Met
                                                           1

gat caa cca atg aaa cca aaa act tgc tct gaa tct gat ttt gct gat       104
Asp Gln Pro Met Lys Pro Lys Thr Cys Ser Glu Ser Asp Phe Ala Asp
          5                  10                  15

gat tcc tct gct tct tct tct tct tct tcg gga caa aat ctc aga gga       152
Asp Ser Ser Ala Ser Ser Ser Ser Ser Ser Gly Gln Asn Leu Arg Gly
        20                  25                  30

gct gag atg gtg gtg gaa gtg aag aag gaa gca gtt tgt tcc cag aaa       200
Ala Glu Met Val Val Glu Val Lys Lys Glu Ala Val Cys Ser Gln Lys
    35                  40                  45

gca gag cga gag aag ctt cgt aga gat aag ctt aag gaa cag ttt ctt       248
Ala Glu Arg Glu Lys Leu Arg Arg Asp Lys Leu Lys Glu Gln Phe Leu
50                  55                  60                  65

gag ctt gga aat gca ctt gat ccg aat agg cct aag agt gac aaa gcc       296
Glu Leu Gly Asn Ala Leu Asp Pro Asn Arg Pro Lys Ser Asp Lys Ala
                70                  75                  80

tca gtt ctc act gat aca ata caa atg ctc aag gat gta atg aac caa       344
Ser Val Leu Thr Asp Thr Ile Gln Met Leu Lys Asp Val Met Asn Gln
            85                  90                  95

gtt gat aga cta aaa gct gag tat gaa aca cta tct caa gag tct cgt       392
Val Asp Arg Leu Lys Ala Glu Tyr Glu Thr Leu Ser Gln Glu Ser Arg
        100                 105                 110

gag cta att caa gag aag agt gag ctg aga gag gag aaa gcg act tta       440
Glu Leu Ile Gln Glu Lys Ser Glu Leu Arg Glu Glu Lys Ala Thr Leu
    115                 120                 125

aag tct gat atc gag att ctt aat gct caa tat cag cat aga atc aaa       488
Lys Ser Asp Ile Glu Ile Leu Asn Ala Gln Tyr Gln His Arg Ile Lys
130                 135                 140                 145

acc atg gtt cca tgg gta cct cat tac agt tat cat atc ccc ttc gta       536
Thr Met Val Pro Trp Val Pro His Tyr Ser Tyr His Ile Pro Phe Val
                150                 155                 160

gcc ata act cag ggt cag tcc agt ttt ata cct tat tca gcc tct gtc       584
Ala Ile Thr Gln Gly Gln Ser Ser Phe Ile Pro Tyr Ser Ala Ser Val
            165                 170                 175

aat cct cta acc gaa caa caa gca tcg gtt cag cag cat tct tct tct       632
Asn Pro Leu Thr Glu Gln Gln Ala Ser Val Gln Gln His Ser Ser Ser
        180                 185                 190

tct gcc gat gct tca atg aaa caa gat tcc aaa atc aag ccg tta gat       680
Ser Ala Asp Ala Ser Met Lys Gln Asp Ser Lys Ile Lys Pro Leu Asp
    195                 200                 205

ttg gat ctg atg atg aac agt aac cat tca ggt caa gga aat gat caa       728
```

MBI-20 Sequence Listing.ST25

```
Leu Asp Leu Met Met Asn Ser Asn His Ser Gly Gln Gly Asn Asp Gln
210                 215                 220                 225

aaa gat gat gtt cgt tta aag ctc gag ctt aaa atc cat gcc tct tct      776
Lys Asp Asp Val Arg Leu Lys Leu Glu Leu Lys Ile His Ala Ser Ser
                230                 235                 240

tta gct caa cag gat gtt tct gga aaa gag aag aaa gta agc ttg aca      824
Leu Ala Gln Gln Asp Val Ser Gly Lys Glu Lys Lys Val Ser Leu Thr
            245                 250                 255

acc act gca agc tca tcg aat agt tac tca tta tct caa gct gtt caa      872
Thr Thr Ala Ser Ser Ser Asn Ser Tyr Ser Leu Ser Gln Ala Val Gln
        260                 265                 270

gat agt tcc ccc ggt acc gta aat gac atg ttg aag cca taa              914
Asp Ser Ser Pro Gly Thr Val Asn Asp Met Leu Lys Pro
    275                 280                 285

accaataaac atattcccct gaacttgtgt ttaataccgt gattgagaag gtaccatgat    974

taaacttgtt gtagattatc cacatgatta acgatgtatt cttatcacaa gcaaataaaa   1034

cacaaaagca tttgcttaaa aaa                                          1057
```

<210> 16
<211> 286
<212> PRT
<213> Arabidopsis thaliana

<400> 16

```
Met Asp Gln Pro Met Lys Pro Lys Thr Cys Ser Glu Ser Asp Phe Ala
1               5                   10                  15

Asp Asp Ser Ser Ala Ser Ser Ser Ser Ser Ser Gly Gln Asn Leu Arg
            20                  25                  30

Gly Ala Glu Met Val Val Glu Val Lys Lys Glu Ala Val Cys Ser Gln
        35                  40                  45

Lys Ala Glu Arg Glu Lys Leu Arg Arg Asp Lys Leu Lys Glu Gln Phe
    50                  55                  60

Leu Glu Leu Gly Asn Ala Leu Asp Pro Asn Arg Pro Lys Ser Asp Lys
65                  70                  75                  80

Ala Ser Val Leu Thr Asp Thr Ile Gln Met Leu Lys Asp Val Met Asn
                85                  90                  95

Gln Val Asp Arg Leu Lys Ala Glu Tyr Glu Thr Leu Ser Gln Glu Ser
            100                 105                 110

Arg Glu Leu Ile Gln Glu Lys Ser Glu Leu Arg Glu Glu Lys Ala Thr
        115                 120                 125

Leu Lys Ser Asp Ile Glu Ile Leu Asn Ala Gln Tyr Gln His Arg Ile
    130                 135                 140

Lys Thr Met Val Pro Trp Val Pro His Tyr Ser Tyr His Ile Pro Phe
145                 150                 155                 160

Val Ala Ile Thr Gln Gly Gln Ser Ser Phe Ile Pro Tyr Ser Ala Ser
                165                 170                 175
```

MBI-20 Sequence Listing.ST25

```
Val Asn Pro Leu Thr Glu Gln Gln Ala Ser Val Gln Gln His Ser Ser
            180                 185                 190

Ser Ser Ala Asp Ala Ser Met Lys Gln Asp Ser Lys Ile Lys Pro Leu
        195                 200                 205

Asp Leu Asp Leu Met Met Asn Ser Asn His Ser Gly Gln Gly Asn Asp
    210                 215                 220

Gln Lys Asp Asp Val Arg Leu Lys Leu Glu Leu Lys Ile His Ala Ser
225                 230                 235                 240

Ser Leu Ala Gln Gln Asp Val Ser Gly Lys Glu Lys Lys Val Ser Leu
                245                 250                 255

Thr Thr Thr Ala Ser Ser Ser Asn Ser Tyr Ser Leu Ser Gln Ala Val
            260                 265                 270

Gln Asp Ser Ser Pro Gly Thr Val Asn Asp Met Leu Lys Pro
        275                 280                 285
```

<210> 17
<211> 1571
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (428)..(1402)
<223> G869

<400> 17

```
aggaacagtg aaaggttcgg ttttttgggt ttcgatctga taatcaacaa gaaaaaaggg     60

tttgatttat gtcggctggg tttgaatcga ctgtgatttt gtctttgatt catatctctt    120

ctccgatttc atcatcatct tccccatcat cgtcgtcttt gaaatcttgt cttctcaacg    180

ctcttcactt ctgctgtaat aagcagaggc ttgttctgga gactccttct ctttccatgc    240

gcttaagacc caaaaggact tgttctagtg ttgaagtctt tgggggtttt cacataaagc    300

agcaaaagtt ttctttttttc atagttcgct gagagttttg agttttgata ccaaaaaagt    360

tttgaccttt tagagtgatt ttttgttctt tctgttttct gggtattttt gaggagtggg    420

tttaaca atg gtt gcg att aga aag gaa cag tct ttg agt ggt gtt agt        469
       Met Val Ala Ile Arg Lys Glu Gln Ser Leu Ser Gly Val Ser
       1               5                   10

agc gag att aag aag aga gct aag aga aac act cta tcg tcc ctt cct      517
Ser Glu Ile Lys Lys Arg Ala Lys Arg Asn Thr Leu Ser Ser Leu Pro
15                  20                  25                  30

caa gaa acc caa cct ttg agg aaa gtc cgt att att gtg aat gat cct      565
Gln Glu Thr Gln Pro Leu Arg Lys Val Arg Ile Ile Val Asn Asp Pro
                35                  40                  45

tat gct act gat gat tcc tct agt gat gag gaa gag ctt aag gtt cct      613
Tyr Ala Thr Asp Asp Ser Ser Ser Asp Glu Glu Glu Leu Lys Val Pro
            50                  55                  60

aag cca agg aaa atg aaa cgt atc gtt cgt gag att aac ttt cct tct      661
Lys Pro Arg Lys Met Lys Arg Ile Val Arg Glu Ile Asn Phe Pro Ser
        65                  70                  75
```

```
atg gaa gtt tct gaa cag cct tct gag agt tct tct cag gac agt act      709
Met Glu Val Ser Glu Gln Pro Ser Glu Ser Ser Ser Gln Asp Ser Thr
    80                  85                  90

aaa act gat ggc aag ata gct gtg tca gct tct cct gct gtt cct agg      757
Lys Thr Asp Gly Lys Ile Ala Val Ser Ala Ser Pro Ala Val Pro Arg
95                  100                 105                 110

aag aag cct gtt ggt gtt agg caa agg aaa tgg ggg aaa tgg gct gct      805
Lys Lys Pro Val Gly Val Arg Gln Arg Lys Trp Gly Lys Trp Ala Ala
                115                 120                 125

gag att aga gat cct att aag aaa act agg act tgg ttg ggt act ttt      853
Glu Ile Arg Asp Pro Ile Lys Lys Thr Arg Thr Trp Leu Gly Thr Phe
            130                 135                 140

gat act ctt gaa gaa gct gct aaa gct tat gat gct aag aag ctt gag      901
Asp Thr Leu Glu Glu Ala Ala Lys Ala Tyr Asp Ala Lys Lys Leu Glu
        145                 150                 155

ttt gat gct att gtt gct gga aat gtg tcc act act aaa cgt gat gtt      949
Phe Asp Ala Ile Val Ala Gly Asn Val Ser Thr Thr Lys Arg Asp Val
    160                 165                 170

tct tca tct gag act agc caa tgc tct cgt tct tca cct gtt gtt cct      997
Ser Ser Ser Glu Thr Ser Gln Cys Ser Arg Ser Ser Pro Val Val Pro
175                 180                 185                 190

gtt gag caa gat gac act tct gca tca gct ctc act tgt gtc aac aac     1045
Val Glu Gln Asp Asp Thr Ser Ala Ser Ala Leu Thr Cys Val Asn Asn
                195                 200                 205

cct gat gac gtc tcg acc gtt gct cca act gct cca act cca aat gtt     1093
Pro Asp Asp Val Ser Thr Val Ala Pro Thr Ala Pro Thr Pro Asn Val
            210                 215                 220

cct gct ggt gga aac aag gaa acg ttg ttc gat ttc gac ttt act aat     1141
Pro Ala Gly Gly Asn Lys Glu Thr Leu Phe Asp Phe Asp Phe Thr Asn
        225                 230                 235

cta cag atc cct gat ttt ggt ttc ttg gca gag gag caa caa gac cta     1189
Leu Gln Ile Pro Asp Phe Gly Phe Leu Ala Glu Glu Gln Gln Asp Leu
    240                 245                 250

gac ttc gat tgt ttc ctc gcg gat gat cag ttt gat gat ttc ggc ttg     1237
Asp Phe Asp Cys Phe Leu Ala Asp Asp Gln Phe Asp Asp Phe Gly Leu
255                 260                 265                 270

ctt gat gac att caa gga ttc gaa gat aac ggt cca agt gcg tta cca     1285
Leu Asp Asp Ile Gln Gly Phe Glu Asp Asn Gly Pro Ser Ala Leu Pro
                275                 280                 285

gat ttc gac ttt gcg gat gtt gaa gat ctt cag cta gct gac tct agt     1333
Asp Phe Asp Phe Ala Asp Val Glu Asp Leu Gln Leu Ala Asp Ser Ser
            290                 295                 300

ttc ggt ttc ctt gat caa ctt gct cct atc aac atc tct tgc cca tta     1381
Phe Gly Phe Leu Asp Gln Leu Ala Pro Ile Asn Ile Ser Cys Pro Leu
        305                 310                 315

aaa agt ttt gca gct tca tag gatcttgctt agtaatgtta agtgagaaga        1432
Lys Ser Phe Ala Ala Ser
    320

gtgttttgtt ttttcgttta tgctttagta atttaagaca tacaaaagtg tgtgttccgg   1492

attgtagtaa gatcttaaga cataaagccg ggttttgcaa ttaggaatcg agttttaatg   1552

aagttttagt ttatgtttg                                               1571
```

<210> 18
<211> 324
<212> PRT

MBI-20 Sequence Listing.ST25

<213> Arabidopsis thaliana

<400> 18

Met Val Ala Ile Arg Lys Glu Gln Ser Leu Ser Gly Val Ser Ser Glu
1 5 10 15

Ile Lys Lys Arg Ala Lys Arg Asn Thr Leu Ser Ser Leu Pro Gln Glu
20 25 30

Thr Gln Pro Leu Arg Lys Val Arg Ile Ile Val Asn Asp Pro Tyr Ala
35 40 45

Thr Asp Asp Ser Ser Ser Asp Glu Glu Glu Leu Lys Val Pro Lys Pro
50 55 60

Arg Lys Met Lys Arg Ile Val Arg Glu Ile Asn Phe Pro Ser Met Glu
65 70 75 80

Val Ser Glu Gln Pro Ser Glu Ser Ser Ser Gln Asp Ser Thr Lys Thr
85 90 95

Asp Gly Lys Ile Ala Val Ser Ala Ser Pro Ala Val Pro Arg Lys Lys
100 105 110

Pro Val Gly Val Arg Gln Arg Lys Trp Gly Lys Trp Ala Ala Glu Ile
115 120 125

Arg Asp Pro Ile Lys Lys Thr Arg Thr Trp Leu Gly Thr Phe Asp Thr
130 135 140

Leu Glu Glu Ala Ala Lys Ala Tyr Asp Ala Lys Lys Leu Glu Phe Asp
145 150 155 160

Ala Ile Val Ala Gly Asn Val Ser Thr Thr Lys Arg Asp Val Ser Ser
165 170 175

Ser Glu Thr Ser Gln Cys Ser Arg Ser Ser Pro Val Val Pro Val Glu
180 185 190

Gln Asp Asp Thr Ser Ala Ser Ala Leu Thr Cys Val Asn Asn Pro Asp
195 200 205

Asp Val Ser Thr Val Ala Pro Thr Ala Pro Thr Pro Asn Val Pro Ala
210 215 220

Gly Gly Asn Lys Glu Thr Leu Phe Asp Phe Asp Phe Thr Asn Leu Gln
225 230 235 240

Ile Pro Asp Phe Gly Phe Leu Ala Glu Glu Gln Gln Asp Leu Asp Phe
245 250 255

Asp Cys Phe Leu Ala Asp Asp Gln Phe Asp Asp Phe Gly Leu Leu Asp
260 265 270

Asp Ile Gln Gly Phe Glu Asp Asn Gly Pro Ser Ala Leu Pro Asp Phe
275 280 285

MBI-20 Sequence Listing.ST25

```
Asp Phe Ala Asp Val Glu Asp Leu Gln Leu Ala Asp Ser Ser Phe Gly
    290                 295                 300

Phe Leu Asp Gln Leu Ala Pro Ile Asn Ile Ser Cys Pro Leu Lys Ser
305                 310                 315                 320

Phe Ala Ala Ser
```

<210> 19
<211> 1322
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (104)..(1084)
<223> G1133

<400> 19

```
ttcaagaaag aatcaccaag tgttgcgttc cacacatttg agcaacagct tccacaatcg     60

tattgtattc ctgtaaagtt cccttggctt aaactgcaag agc atg cct ctt gat      115
                                                Met Pro Leu Asp
                                                1

acc aaa cag cag aaa tgg ttg cca tta ggc tta aat cct caa gct tgt      163
Thr Lys Gln Gln Lys Trp Leu Pro Leu Gly Leu Asn Pro Gln Ala Cys
5                   10                  15                  20

gtc cag gac aag gcg act gag tat ttc cgt cct gga att cct ttt ccg      211
Val Gln Asp Lys Ala Thr Glu Tyr Phe Arg Pro Gly Ile Pro Phe Pro
                25                  30                  35

gaa ctc ggt aaa gtt tat gca gct gag cat cag ttt cgc tat ttg cag      259
Glu Leu Gly Lys Val Tyr Ala Ala Glu His Gln Phe Arg Tyr Leu Gln
            40                  45                  50

cca ccg ttc caa gcc tta ttg tct aga tat gat cag cag tct tgt gga      307
Pro Pro Phe Gln Ala Leu Leu Ser Arg Tyr Asp Gln Gln Ser Cys Gly
        55                  60                  65

aaa caa gtt tca tgt ttg aat ggg cga tct agc aac ggt gct gct cca      355
Lys Gln Val Ser Cys Leu Asn Gly Arg Ser Ser Asn Gly Ala Ala Pro
    70                  75                  80

gag ggg gca ctc aag tct tct cgg aaa aga ttt ata gta ttc gat cag      403
Glu Gly Ala Leu Lys Ser Ser Arg Lys Arg Phe Ile Val Phe Asp Gln
85                  90                  95                  100

tcg gga gag cag act cgt ttg tta caa tgt gga ttt cct ctg cgg ttt      451
Ser Gly Glu Gln Thr Arg Leu Leu Gln Cys Gly Phe Pro Leu Arg Phe
                105                 110                 115

cct tct tct atg gat gca gag cga ggg aac att ctc ggt gcc cta cac      499
Pro Ser Ser Met Asp Ala Glu Arg Gly Asn Ile Leu Gly Ala Leu His
            120                 125                 130

cca gag aaa ggg ttt agt aaa gat cat gcc att caa gaa aag ata ttg      547
Pro Glu Lys Gly Phe Ser Lys Asp His Ala Ile Gln Glu Lys Ile Leu
        135                 140                 145

caa cat gaa gat cat gaa aat ggc gaa gaa gac tcg gaa atg cac gaa      595
Gln His Glu Asp His Glu Asn Gly Glu Glu Asp Ser Glu Met His Glu
    150                 155                 160

gac act gag gaa atc aac gcg tta ctg tat tct gat gat gac gat aat      643
Asp Thr Glu Glu Ile Asn Ala Leu Leu Tyr Ser Asp Asp Asp Asp Asn
165                 170                 175                 180
```

```
gat gat tgg gaa agt gat gat gaa gta atg agc act ggt cac tct cca      691
Asp Asp Trp Glu Ser Asp Asp Glu Val Met Ser Thr Gly His Ser Pro
                185                 190                 195

ttc aca gtt gaa caa caa gcg tgc aac ata aca aca gaa gag ctg gat      739
Phe Thr Val Glu Gln Gln Ala Cys Asn Ile Thr Thr Glu Glu Leu Asp
            200                 205                 210

gaa act gaa agc act gtt gat ggt cca ctt ctt aaa aga cag aaa cta      787
Glu Thr Glu Ser Thr Val Asp Gly Pro Leu Leu Lys Arg Gln Lys Leu
        215                 220                 225

ctg gac cat tcg tac aga gac tca tca cca tcc ctt gtg ggc acc act      835
Leu Asp His Ser Tyr Arg Asp Ser Ser Pro Ser Leu Val Gly Thr Thr
    230                 235                 240

aaa gtc aaa ggc tta tca gat gaa aac ctt cct gaa tca aac att tca      883
Lys Val Lys Gly Leu Ser Asp Glu Asn Leu Pro Glu Ser Asn Ile Ser
245                 250                 255                 260

agc aaa caa gaa acg ggt tct ggt ttg agc gac gag cag tca aga aaa      931
Ser Lys Gln Glu Thr Gly Ser Gly Leu Ser Asp Glu Gln Ser Arg Lys
                265                 270                 275

gac aag att cac acc gct ctg aga atc ctg gag agt gta gtt cca ggg      979
Asp Lys Ile His Thr Ala Leu Arg Ile Leu Glu Ser Val Val Pro Gly
            280                 285                 290

gca aag gga aaa gaa gct ctt tta cta cta gac gaa gcc att gat tac     1027
Ala Lys Gly Lys Glu Ala Leu Leu Leu Leu Asp Glu Ala Ile Asp Tyr
        295                 300                 305

ctc aag ttg ctg aag caa agc tta aac tca tca aag ggt ttg aat aac     1075
Leu Lys Leu Leu Lys Gln Ser Leu Asn Ser Ser Lys Gly Leu Asn Asn
    310                 315                 320

cat tgg tga aaaacctaca acccctttg tcctattgat aaggcatgtt              1124
His Trp
325

tggttggtta aagagaagac atgggacaaa agataatcaa tgaggtaaag gactgatgaa   1184

gaagattctc tcaaattcat taacgtgggt ttgaaacaat tagaacacgc ctggtgaccc   1244

tagtgggacc gtatccactg ttcatctagc tggatcaata gtggtttact tttggatttg   1304

gcatgctctc tcaaaaaa                                                1322
```

<210> 20
<211> 326
<212> PRT
<213> Arabidopsis thaliana

<400> 20

```
Met Pro Leu Asp Thr Lys Gln Gln Lys Trp Leu Pro Leu Gly Leu Asn
1               5                   10                  15

Pro Gln Ala Cys Val Gln Asp Lys Ala Thr Glu Tyr Phe Arg Pro Gly
            20                  25                  30

Ile Pro Phe Pro Glu Leu Gly Lys Val Tyr Ala Ala Glu His Gln Phe
        35                  40                  45

Arg Tyr Leu Gln Pro Pro Phe Gln Ala Leu Leu Ser Arg Tyr Asp Gln
    50                  55                  60

Gln Ser Cys Gly Lys Gln Val Ser Cys Leu Asn Gly Arg Ser Ser Asn
65                  70                  75                  80
```

MBI-20 Sequence Listing.ST25

Gly Ala Ala Pro Glu Gly Ala Leu Lys Ser Ser Arg Lys Arg Phe Ile
85 90 95

Val Phe Asp Gln Ser Gly Glu Gln Thr Arg Leu Leu Gln Cys Gly Phe
100 105 110

Pro Leu Arg Phe Pro Ser Ser Met Asp Ala Glu Arg Gly Asn Ile Leu
115 120 125

Gly Ala Leu His Pro Glu Lys Gly Phe Ser Lys Asp His Ala Ile Gln
130 135 140

Glu Lys Ile Leu Gln His Glu Asp His Glu Asn Gly Glu Glu Asp Ser
145 150 155 160

Glu Met His Glu Asp Thr Glu Glu Ile Asn Ala Leu Leu Tyr Ser Asp
165 170 175

Asp Asp Asp Asn Asp Asp Trp Glu Ser Asp Asp Glu Val Met Ser Thr
180 185 190

Gly His Ser Pro Phe Thr Val Glu Gln Gln Ala Cys Asn Ile Thr Thr
195 200 205

Glu Glu Leu Asp Glu Thr Glu Ser Thr Val Asp Gly Pro Leu Leu Lys
210 215 220

Arg Gln Lys Leu Leu Asp His Ser Tyr Arg Asp Ser Ser Pro Ser Leu
225 230 235 240

Val Gly Thr Thr Lys Val Lys Gly Leu Ser Asp Glu Asn Leu Pro Glu
245 250 255

Ser Asn Ile Ser Ser Lys Gln Glu Thr Gly Ser Gly Leu Ser Asp Glu
260 265 270

Gln Ser Arg Lys Asp Lys Ile His Thr Ala Leu Arg Ile Leu Glu Ser
275 280 285

Val Val Pro Gly Ala Lys Gly Lys Glu Ala Leu Leu Leu Leu Asp Glu
290 295 300

Ala Ile Asp Tyr Leu Lys Leu Leu Lys Gln Ser Leu Asn Ser Ser Lys
305 310 315 320

Gly Leu Asn Asn His Trp
325

<210> 21
<211> 859
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (62)..(718)
<223> G1266

<400> 21
caatccacta acgatcccta accgaaaaca gagtagtcaa gaaacagagt atttttttcta 60

c atg gat cca ttt tta att cag tcc cca ttc tcc ggc ttc tca ccg gaa 109
Met Asp Pro Phe Leu Ile Gln Ser Pro Phe Ser Gly Phe Ser Pro Glu
1 5 10 15

tat tct atc gga tct tct cca gat tct ttc tca tcc tct tct tct aac 157
Tyr Ser Ile Gly Ser Ser Pro Asp Ser Phe Ser Ser Ser Ser Ser Asn
20 25 30

aat tac tct ctt ccc ttc aac gag aac gac tca gag gaa atg ttt ctc 205
Asn Tyr Ser Leu Pro Phe Asn Glu Asn Asp Ser Glu Glu Met Phe Leu
35 40 45

tac ggt cta atc gag cag tcc acg caa caa acc tat att gac tcg gat 253
Tyr Gly Leu Ile Glu Gln Ser Thr Gln Gln Thr Tyr Ile Asp Ser Asp
50 55 60

agt caa gac ctt ccg atc aaa tcc gta agc tca aga aag tca gag aag 301
Ser Gln Asp Leu Pro Ile Lys Ser Val Ser Ser Arg Lys Ser Glu Lys
65 70 75 80

tct tac aga ggc gta aga cga cgg cca tgg ggg aaa ttc gcg gcg gag 349
Ser Tyr Arg Gly Val Arg Arg Arg Pro Trp Gly Lys Phe Ala Ala Glu
85 90 95

ata aga gat tcg act aga aac ggt att agg gtt tgg ctc ggg acg ttc 397
Ile Arg Asp Ser Thr Arg Asn Gly Ile Arg Val Trp Leu Gly Thr Phe
100 105 110

gaa agc gcg gaa gag gcg gct tta gcc tac gat caa gct gct ttc tcg 445
Glu Ser Ala Glu Glu Ala Ala Leu Ala Tyr Asp Gln Ala Ala Phe Ser
115 120 125

atg aga ggg tcc tcg gcg att ctc aat ttt tcg gcg gag aga gtt caa 493
Met Arg Gly Ser Ser Ala Ile Leu Asn Phe Ser Ala Glu Arg Val Gln
130 135 140

gag tcg ctt tcg gag att aaa tat acc tac gag gat ggt tgt tct ccg 541
Glu Ser Leu Ser Glu Ile Lys Tyr Thr Tyr Glu Asp Gly Cys Ser Pro
145 150 155 160

gtt gtg gcg ttg aag agg aaa cac tcg atg aga cgg aga atg acc aat 589
Val Val Ala Leu Lys Arg Lys His Ser Met Arg Arg Arg Met Thr Asn
165 170 175

aag aag acg aaa gat agt gac ttt gat cac cgc tcc gtg aag tta gat 637
Lys Lys Thr Lys Asp Ser Asp Phe Asp His Arg Ser Val Lys Leu Asp
180 185 190

aat gta gtt gtc ttt gag gat ttg gga gaa cag tac ctt gag gag ctt 685
Asn Val Val Val Phe Glu Asp Leu Gly Glu Gln Tyr Leu Glu Glu Leu
195 200 205

ttg ggg tct tct gaa aat agt ggg act tgg tga aagattagga tttgtattag 738
Leu Gly Ser Ser Glu Asn Ser Gly Thr Trp
210 215

ggaccttaag tttgaagtgg ttgattaatt ttaaccctaa tatgtttttt gtttgcttaa 798

atatttgatt ctattgagaa acatcgaaaa cagtttgtat gtacttttgt gatacttggc 858

g 859

<210> 22
<211> 218
<212> PRT
<213> Arabidopsis thaliana

<400> 22

Met Asp Pro Phe Leu Ile Gln Ser Pro Phe Ser Gly Phe Ser Pro Glu

MBI-20 Sequence Listing.ST25

1 5 10 15

Tyr Ser Ile Gly Ser Ser Pro Asp Ser Phe Ser Ser Ser Ser Ser Asn
20 25 30

Asn Tyr Ser Leu Pro Phe Asn Glu Asn Asp Ser Glu Glu Met Phe Leu
35 40 45

Tyr Gly Leu Ile Glu Gln Ser Thr Gln Gln Thr Tyr Ile Asp Ser Asp
50 55 60

Ser Gln Asp Leu Pro Ile Lys Ser Val Ser Ser Arg Lys Ser Glu Lys
65 70 75 80

Ser Tyr Arg Gly Val Arg Arg Arg Pro Trp Gly Lys Phe Ala Ala Glu
85 90 95

Ile Arg Asp Ser Thr Arg Asn Gly Ile Arg Val Trp Leu Gly Thr Phe
100 105 110

Glu Ser Ala Glu Glu Ala Ala Leu Ala Tyr Asp Gln Ala Ala Phe Ser
115 120 125

Met Arg Gly Ser Ser Ala Ile Leu Asn Phe Ser Ala Glu Arg Val Gln
130 135 140

Glu Ser Leu Ser Glu Ile Lys Tyr Thr Tyr Glu Asp Gly Cys Ser Pro
145 150 155 160

Val Val Ala Leu Lys Arg Lys His Ser Met Arg Arg Arg Met Thr Asn
165 170 175

Lys Lys Thr Lys Asp Ser Asp Phe Asp His Arg Ser Val Lys Leu Asp
180 185 190

Asn Val Val Val Phe Glu Asp Leu Gly Glu Gln Tyr Leu Glu Glu Leu
195 200 205

Leu Gly Ser Ser Glu Asn Ser Gly Thr Trp
210 215

<210> 23
<211> 1137
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (54)..(914)
<223> G1324

<400> 23
cgaaaacacc acaaaccaaa tatcattaag taattaggaa acttaaacta agt atg 56
Met
1

gaa aat tcg atg aag aag aag aag agc ttc aaa gaa agt gaa gat gaa 104
Glu Asn Ser Met Lys Lys Lys Lys Ser Phe Lys Glu Ser Glu Asp Glu
5 10 15

MBI-20 Sequence Listing.ST25

```
gaa cta aga aga ggg cct tgg act ttg gag gaa gac aca ctt ctc aca      152
Glu Leu Arg Arg Gly Pro Trp Thr Leu Glu Glu Asp Thr Leu Leu Thr
        20                  25                  30

aat tac atc ctc cat aac ggt gag ggt cgt tgg aat cac gtc gcc aaa      200
Asn Tyr Ile Leu His Asn Gly Glu Gly Arg Trp Asn His Val Ala Lys
    35                  40                  45

tgt gct ggg cta aag aga act ggg aaa agt tgt aga ttg aga tgg ttg      248
Cys Ala Gly Leu Lys Arg Thr Gly Lys Ser Cys Arg Leu Arg Trp Leu
50                  55                  60                  65

aat tac ttg aaa ccc gac ata aga cga ggg aat ctt act cct caa gaa      296
Asn Tyr Leu Lys Pro Asp Ile Arg Arg Gly Asn Leu Thr Pro Gln Glu
                70                  75                  80

cag ctt ttg atc ctt gag ctt cac tct aaa tgg ggt aat agg tgg tcc      344
Gln Leu Leu Ile Leu Glu Leu His Ser Lys Trp Gly Asn Arg Trp Ser
            85                  90                  95

aag att gca cag tac ttg cca gga aga acg gat aac gag atc aag aac      392
Lys Ile Ala Gln Tyr Leu Pro Gly Arg Thr Asp Asn Glu Ile Lys Asn
        100                 105                 110

tat tgg aga aca aga gtt caa aaa caa gct cgt caa ctc aac atc gaa      440
Tyr Trp Arg Thr Arg Val Gln Lys Gln Ala Arg Gln Leu Asn Ile Glu
    115                 120                 125

tct aac agc gac aag ttc ttt gac gct gtt cgt agt ttt tgg gtc cct      488
Ser Asn Ser Asp Lys Phe Phe Asp Ala Val Arg Ser Phe Trp Val Pro
130                 135                 140                 145

aga ttg atc gag aag atg gaa caa aac tca tcc act act act act tat      536
Arg Leu Ile Glu Lys Met Glu Gln Asn Ser Ser Thr Thr Thr Thr Tyr
                150                 155                 160

tgt tgt ccc caa aac aac aac aac aac tct ctt ctt ctt cct tct caa      584
Cys Cys Pro Gln Asn Asn Asn Asn Asn Ser Leu Leu Leu Pro Ser Gln
            165                 170                 175

tct cac gac tct tta agt atg caa aaa gat ata gat tac tcg ggt ttc      632
Ser His Asp Ser Leu Ser Met Gln Lys Asp Ile Asp Tyr Ser Gly Phe
        180                 185                 190

agc aac ata gac ggt tct tct tca act tct act tgc atg tct cat cta      680
Ser Asn Ile Asp Gly Ser Ser Ser Thr Ser Thr Cys Met Ser His Leu
    195                 200                 205

aca aca gtt cca cac ttt atg gat caa agc aac acc aat atc atc gat      728
Thr Thr Val Pro His Phe Met Asp Gln Ser Asn Thr Asn Ile Ile Asp
210                 215                 220                 225

ggc tcg atg tgt ttc cat gaa ggc aat gtt caa gaa ttc gga gga tat      776
Gly Ser Met Cys Phe His Glu Gly Asn Val Gln Glu Phe Gly Gly Tyr
                230                 235                 240

gtt cct ggc atg gag gat tac atg gta aac tcg gac atc tca atg gaa      824
Val Pro Gly Met Glu Asp Tyr Met Val Asn Ser Asp Ile Ser Met Glu
            245                 250                 255

tgt cac gtg gcg gat ggt tat tca gcg tac gag gat gtt aca caa gat      872
Cys His Val Ala Asp Gly Tyr Ser Ala Tyr Glu Asp Val Thr Gln Asp
        260                 265                 270

ccc atg tgg aat gtg gat gac att tgg cag ttt agg gag taa              914
Pro Met Trp Asn Val Asp Asp Ile Trp Gln Phe Arg Glu
    275                 280                 285

ttaagtcgtc aagagatgag atggtagagc ctaccactac ggttctatta tatggactaa    974

tatacttctt ttgcttaact aagcaaaaag tttcgaacct tttacccata ttatctcggg   1034

ttggagacta gaacatgtta aatttgtatc ttctttgttg cgagtactta ctaagtcatt   1094

ggataaatat ttataatgat agtttcttgt acaaaaaaaa aaa                     1137
```

MBI-20 Sequence Listing.ST25

<210> 24
<211> 286
<212> PRT
<213> Arabidopsis thaliana

<400> 24

Met Glu Asn Ser Met Lys Lys Lys Lys Ser Phe Lys Glu Ser Glu Asp
1 5 10 15

Glu Glu Leu Arg Arg Gly Pro Trp Thr Leu Glu Glu Asp Thr Leu Leu
20 25 30

Thr Asn Tyr Ile Leu His Asn Gly Glu Gly Arg Trp Asn His Val Ala
35 40 45

Lys Cys Ala Gly Leu Lys Arg Thr Gly Lys Ser Cys Arg Leu Arg Trp
50 55 60

Leu Asn Tyr Leu Lys Pro Asp Ile Arg Arg Gly Asn Leu Thr Pro Gln
65 70 75 80

Glu Gln Leu Leu Ile Leu Glu Leu His Ser Lys Trp Gly Asn Arg Trp
85 90 95

Ser Lys Ile Ala Gln Tyr Leu Pro Gly Arg Thr Asp Asn Glu Ile Lys
100 105 110

Asn Tyr Trp Arg Thr Arg Val Gln Lys Gln Ala Arg Gln Leu Asn Ile
115 120 125

Glu Ser Asn Ser Asp Lys Phe Phe Asp Ala Val Arg Ser Phe Trp Val
130 135 140

Pro Arg Leu Ile Glu Lys Met Glu Gln Asn Ser Ser Thr Thr Thr Thr
145 150 155 160

Tyr Cys Cys Pro Gln Asn Asn Asn Asn Asn Ser Leu Leu Leu Pro Ser
165 170 175

Gln Ser His Asp Ser Leu Ser Met Gln Lys Asp Ile Asp Tyr Ser Gly
180 185 190

Phe Ser Asn Ile Asp Gly Ser Ser Ser Thr Ser Thr Cys Met Ser His
195 200 205

Leu Thr Thr Val Pro His Phe Met Asp Gln Ser Asn Thr Asn Ile Ile
210 215 220

Asp Gly Ser Met Cys Phe His Glu Gly Asn Val Gln Glu Phe Gly Gly
225 230 235 240

Tyr Val Pro Gly Met Glu Asp Tyr Met Val Asn Ser Asp Ile Ser Met
245 250 255

Glu Cys His Val Ala Asp Gly Tyr Ser Ala Tyr Glu Asp Val Thr Gln
260 265 270

MBI-20 Sequence Listing.ST25

```
Asp Pro Met Trp Asn Val Asp Asp Ile Trp Gln Phe Arg Glu
        275                 280                 285


<210>  25
<211>  1630
<212>  DNA
<213>  Arabidopsis thaliana

<220>
<221>  CDS
<222>  (97)..(1398)
<223>  G1337


<400>  25
aatggatttg tcatcattct tctcaccgtc cttagtctct gaaaataaat tctgattttg     60

atttcgaatt ttagggattt tgagagagag tcagtt atg agt agt tcg gag aga      114
                                       Met Ser Ser Ser Glu Arg
                                       1               5

gta ccg tgc gat ttc tgc ggc gag cgt acg gcg gtt ttg ttt tgt aga      162
Val Pro Cys Asp Phe Cys Gly Glu Arg Thr Ala Val Leu Phe Cys Arg
            10                  15                  20

gcc gat acg gcg aag ctg tgt ttg cct tgt gat cag caa gtt cac acg      210
Ala Asp Thr Ala Lys Leu Cys Leu Pro Cys Asp Gln Gln Val His Thr
        25                  30                  35

gcg aat ctg ttg tcg agg aag cac gtg cga tct cag atc tgc gat aat      258
Ala Asn Leu Leu Ser Arg Lys His Val Arg Ser Gln Ile Cys Asp Asn
    40                  45                  50

tgc ggt aac gag cca gtc tct gtt cgg tgt ttc acc gat aat ctg att      306
Cys Gly Asn Glu Pro Val Ser Val Arg Cys Phe Thr Asp Asn Leu Ile
55                  60                  65                  70

ttg tgt cag gag tgt gat tgg gat gtt cac gga agt tgt tca gtt tcc      354
Leu Cys Gln Glu Cys Asp Trp Asp Val His Gly Ser Cys Ser Val Ser
                75                  80                  85

gat gct cat gtt cga tcc gcc gtg gaa ggt ttt tcc ggt tgt cca tcg      402
Asp Ala His Val Arg Ser Ala Val Glu Gly Phe Ser Gly Cys Pro Ser
            90                  95                  100

gcg ttg gag ctt gct gct tta tgg gga ctt gat ttg gag caa ggg agg      450
Ala Leu Glu Leu Ala Ala Leu Trp Gly Leu Asp Leu Glu Gln Gly Arg
        105                 110                 115

aaa gat gaa gag aat caa gtt ccg atg atg gcg atg atg atg gat aat      498
Lys Asp Glu Glu Asn Gln Val Pro Met Met Ala Met Met Met Asp Asn
    120                 125                 130

ttc ggg atg cag ttg gat tct tgg gtt ttg gga tct aat gaa ttg att      546
Phe Gly Met Gln Leu Asp Ser Trp Val Leu Gly Ser Asn Glu Leu Ile
135                 140                 145                 150

gtt ccc agc gat acg acg ttt aag aag cgt gga tct tgt gga tct agt      594
Val Pro Ser Asp Thr Thr Phe Lys Lys Arg Gly Ser Cys Gly Ser Ser
                155                 160                 165

tgt ggg agg tat aag cag gta ttg tgt aag cag ctt gag gag ttg ctt      642
Cys Gly Arg Tyr Lys Gln Val Leu Cys Lys Gln Leu Glu Glu Leu Leu
            170                 175                 180

aag agt ggt gtt gtc ggt ggt gat ggc gat gat ggt gat cgt gac cgt      690
Lys Ser Gly Val Val Gly Gly Asp Gly Asp Asp Gly Asp Arg Asp Arg
        185                 190                 195

gat tgt gac cgt gag ggt gct tgt gat gga gat gga gat gga gaa gca      738
Asp Cys Asp Arg Glu Gly Ala Cys Asp Gly Asp Gly Asp Gly Glu Ala
    200                 205                 210
```

MBI-20 Sequence Listing.ST25

```
gga gag ggg ctt atg gtt ccg gag atg tca gag aga ttg aaa tgg tca      786
Gly Glu Gly Leu Met Val Pro Glu Met Ser Glu Arg Leu Lys Trp Ser
215                 220                 225                 230

aga gat gtt gag gag atc aat ggt ggc gga gga gga gga gtt aac cag      834
Arg Asp Val Glu Glu Ile Asn Gly Gly Gly Gly Gly Gly Val Asn Gln
                235                 240                 245

cag tgg aat gct act act act aat cct agt ggt ggc cag agt tct cag      882
Gln Trp Asn Ala Thr Thr Thr Asn Pro Ser Gly Gly Gln Ser Ser Gln
            250                 255                 260

ata tgg gat ttt aac ttg gga cag tca cgg gga cct gag gat acg agt      930
Ile Trp Asp Phe Asn Leu Gly Gln Ser Arg Gly Pro Glu Asp Thr Ser
        265                 270                 275

cga gtg gaa gct gca tat gta ggg aaa ggt gct gct tct tca ttc aca      978
Arg Val Glu Ala Ala Tyr Val Gly Lys Gly Ala Ala Ser Ser Phe Thr
    280                 285                 290

atc aac aat ttt gtt gac cat atg aat gaa act tgt tcc act aat gtg     1026
Ile Asn Asn Phe Val Asp His Met Asn Glu Thr Cys Ser Thr Asn Val
295                 300                 305                 310

aaa ggt gtc aaa gag att aaa aag gat gac tac aag cga tca act tca     1074
Lys Gly Val Lys Glu Ile Lys Lys Asp Asp Tyr Lys Arg Ser Thr Ser
                315                 320                 325

ggc cag gta caa cca aca aaa tct gag agc aac aat cgt cca att acc     1122
Gly Gln Val Gln Pro Thr Lys Ser Glu Ser Asn Asn Arg Pro Ile Thr
            330                 335                 340

ttt ggc tct gag aaa ggt tcg aac tcc tcc agt gac ttg cat ttc aca     1170
Phe Gly Ser Glu Lys Gly Ser Asn Ser Ser Ser Asp Leu His Phe Thr
        345                 350                 355

gag cat att gct gga act agt tgt aag acc aca aga cta gtt gca act     1218
Glu His Ile Ala Gly Thr Ser Cys Lys Thr Thr Arg Leu Val Ala Thr
    360                 365                 370

aag gct gat ctg gag cgg ctg gct cag aac aga gga gat gca atg cag     1266
Lys Ala Asp Leu Glu Arg Leu Ala Gln Asn Arg Gly Asp Ala Met Gln
375                 380                 385                 390

cgt tac aag gaa aag agg aag aca cgg aga tat gat aag acc ata agg     1314
Arg Tyr Lys Glu Lys Arg Lys Thr Arg Arg Tyr Asp Lys Thr Ile Arg
                395                 400                 405

tat gaa tcg agg aag gca aga gct gac act agg ttg cgt gtc aga ggc     1362
Tyr Glu Ser Arg Lys Ala Arg Ala Asp Thr Arg Leu Arg Val Arg Gly
            410                 415                 420

aga ttt gtg aaa gct agt gaa gct cct tac cct taa ccttaagttt          1408
Arg Phe Val Lys Ala Ser Glu Ala Pro Tyr Pro
        425                 430

tttcacatag gcttcctttt agctacaaac ttagttactt tttttactcc actgcctcat   1468

aaatgtacag accggtctcg tttcatctgg ccgcccttct tgttttattg ccttatctgg   1528

ccctttttatg taccttggaa tcttatctag tttaaaaaag attgtaacct tctagaaaac   1588

catattctgt tgacagtata tacatgtcta tccaagcaaa aa                      1630
```

<210> 26
<211> 433
<212> PRT
<213> Arabidopsis thaliana

<400> 26

```
Met Ser Ser Ser Glu Arg Val Pro Cys Asp Phe Cys Gly Glu Arg Thr
1               5                   10                  15
```

MBI-20 Sequence Listing.ST25

Ala Val Leu Phe Cys Arg Ala Asp Thr Ala Lys Leu Cys Leu Pro Cys
20 25 30

Asp Gln Gln Val His Thr Ala Asn Leu Leu Ser Arg Lys His Val Arg
35 40 45

Ser Gln Ile Cys Asp Asn Cys Gly Asn Glu Pro Val Ser Val Arg Cys
50 55 60

Phe Thr Asp Asn Leu Ile Leu Cys Gln Glu Cys Asp Trp Asp Val His
65 70 75 80

Gly Ser Cys Ser Val Ser Asp Ala His Val Arg Ser Ala Val Glu Gly
85 90 95

Phe Ser Gly Cys Pro Ser Ala Leu Glu Leu Ala Ala Leu Trp Gly Leu
100 105 110

Asp Leu Glu Gln Gly Arg Lys Asp Glu Glu Asn Gln Val Pro Met Met
115 120 125

Ala Met Met Met Asp Asn Phe Gly Met Gln Leu Asp Ser Trp Val Leu
130 135 140

Gly Ser Asn Glu Leu Ile Val Pro Ser Asp Thr Thr Phe Lys Lys Arg
145 150 155 160

Gly Ser Cys Gly Ser Ser Cys Gly Arg Tyr Lys Gln Val Leu Cys Lys
165 170 175

Gln Leu Glu Glu Leu Leu Lys Ser Gly Val Val Gly Gly Asp Gly Asp
180 185 190

Asp Gly Asp Arg Asp Arg Asp Cys Asp Arg Glu Gly Ala Cys Asp Gly
195 200 205

Asp Gly Asp Gly Glu Ala Gly Glu Gly Leu Met Val Pro Glu Met Ser
210 215 220

Glu Arg Leu Lys Trp Ser Arg Asp Val Glu Glu Ile Asn Gly Gly Gly
225 230 235 240

Gly Gly Gly Val Asn Gln Gln Trp Asn Ala Thr Thr Thr Asn Pro Ser
245 250 255

Gly Gly Gln Ser Ser Gln Ile Trp Asp Phe Asn Leu Gly Gln Ser Arg
260 265 270

Gly Pro Glu Asp Thr Ser Arg Val Glu Ala Ala Tyr Val Gly Lys Gly
275 280 285

Ala Ala Ser Ser Phe Thr Ile Asn Asn Phe Val Asp His Met Asn Glu
290 295 300

Thr Cys Ser Thr Asn Val Lys Gly Val Lys Glu Ile Lys Lys Asp Asp

MBI-20 Sequence Listing.ST25

305 310 315 320

Tyr Lys Arg Ser Thr Ser Gly Gln Val Gln Pro Thr Lys Ser Glu Ser
325 330 335

Asn Asn Arg Pro Ile Thr Phe Gly Ser Glu Lys Gly Ser Asn Ser Ser
340 345 350

Ser Asp Leu His Phe Thr Glu His Ile Ala Gly Thr Ser Cys Lys Thr
355 360 365

Thr Arg Leu Val Ala Thr Lys Ala Asp Leu Glu Arg Leu Ala Gln Asn
370 375 380

Arg Gly Asp Ala Met Gln Arg Tyr Lys Glu Lys Arg Lys Thr Arg Arg
385 390 395 400

Tyr Asp Lys Thr Ile Arg Tyr Glu Ser Arg Lys Ala Arg Ala Asp Thr
405 410 415

Arg Leu Arg Val Arg Gly Arg Phe Val Lys Ala Ser Glu Ala Pro Tyr
420 425 430

Pro

<210> 27
<211> 768
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (58)..(657)
<223> G975

<400> 27
attactcatc atcaagttcc tactttctct ctgacaaaca tcacagagta agtaaga 57

atg gta cag acg aag aag ttc aga ggt gtc agg caa cgc cat tgg ggt 105
Met Val Gln Thr Lys Lys Phe Arg Gly Val Arg Gln Arg His Trp Gly
1 5 10 15

tct tgg gtc gct gag att cgt cat cct ctc ttg aaa cgg agg att tgg 153
Ser Trp Val Ala Glu Ile Arg His Pro Leu Leu Lys Arg Arg Ile Trp
20 25 30

cta ggg acg ttc gag acc gca gag gag gca gca aga gca tac gac gag 201
Leu Gly Thr Phe Glu Thr Ala Glu Glu Ala Ala Arg Ala Tyr Asp Glu
35 40 45

gcc gcc gtt tta atg agc ggc cgc aac gcc aaa acc aac ttt ccc ctc 249
Ala Ala Val Leu Met Ser Gly Arg Asn Ala Lys Thr Asn Phe Pro Leu
50 55 60

aac aac aac aac acc gga gaa act tcc gag ggc aaa acc gat att tca 297
Asn Asn Asn Asn Thr Gly Glu Thr Ser Glu Gly Lys Thr Asp Ile Ser
65 70 75 80

gct tcg tcc aca atg tca tcc tca aca tca tct tca tcg ctc tct tcc 345
Ala Ser Ser Thr Met Ser Ser Ser Thr Ser Ser Ser Ser Leu Ser Ser
85 90 95

atc ctc agc gcc aaa ctg agg aaa tgc tgc aag tct cct tcc cca tcc 393
Ile Leu Ser Ala Lys Leu Arg Lys Cys Cys Lys Ser Pro Ser Pro Ser

```
            100                 105                 110

ctc acc tgc ctc cgt ctt gac aca gcc agc tcc cat atc ggc gtc tgg      441
Leu Thr Cys Leu Arg Leu Asp Thr Ala Ser Ser His Ile Gly Val Trp
        115                 120                 125

cag aaa cgg gcc ggt tca aag tct gac tcc agc tgg gtc atg acg gtg      489
Gln Lys Arg Ala Gly Ser Lys Ser Asp Ser Ser Trp Val Met Thr Val
    130                 135                 140

gag cta ggt ccc gca agc tcc tcc caa gag act act agt aaa gct tca      537
Glu Leu Gly Pro Ala Ser Ser Ser Gln Glu Thr Thr Ser Lys Ala Ser
145                 150                 155                 160

caa gac gct att ctt gct ccg acc act gaa gtt gaa att ggt ggc agc      585
Gln Asp Ala Ile Leu Ala Pro Thr Thr Glu Val Glu Ile Gly Gly Ser
                165                 170                 175

aga gaa gaa gta ttg gat gag gaa gaa aag gtt gct ttg caa atg ata      633
Arg Glu Glu Val Leu Asp Glu Glu Glu Lys Val Ala Leu Gln Met Ile
            180                 185                 190

gag gag ctt ctc aat aca aac taa atcttatttg cttatatata tgtacctatt     687
Glu Glu Leu Leu Asn Thr Asn
        195

ttcattgctg atttacagcc aaaataatca attataccgt gtattttata gatgttttat    747

attaaaaggt tgttagatat a                                              768


<210>  28
<211>  199
<212>  PRT
<213>  Arabidopsis thaliana

<400>  28

Met Val Gln Thr Lys Lys Phe Arg Gly Val Arg Gln Arg His Trp Gly
1               5                   10                  15

Ser Trp Val Ala Glu Ile Arg His Pro Leu Leu Lys Arg Arg Ile Trp
            20                  25                  30

Leu Gly Thr Phe Glu Thr Ala Glu Glu Ala Ala Arg Ala Tyr Asp Glu
        35                  40                  45

Ala Ala Val Leu Met Ser Gly Arg Asn Ala Lys Thr Asn Phe Pro Leu
    50                  55                  60

Asn Asn Asn Asn Thr Gly Glu Thr Ser Glu Gly Lys Thr Asp Ile Ser
65                  70                  75                  80

Ala Ser Ser Thr Met Ser Ser Ser Thr Ser Ser Ser Ser Leu Ser Ser
                85                  90                  95

Ile Leu Ser Ala Lys Leu Arg Lys Cys Cys Lys Ser Pro Ser Pro Ser
            100                 105                 110

Leu Thr Cys Leu Arg Leu Asp Thr Ala Ser Ser His Ile Gly Val Trp
        115                 120                 125

Gln Lys Arg Ala Gly Ser Lys Ser Asp Ser Ser Trp Val Met Thr Val
    130                 135                 140

Glu Leu Gly Pro Ala Ser Ser Ser Gln Glu Thr Thr Ser Lys Ala Ser
```

```
145                 150                 155                 160

Gln Asp Ala Ile Leu Ala Pro Thr Thr Glu Val Glu Ile Gly Gly Ser
                165                 170                 175

Arg Glu Glu Val Leu Asp Glu Glu Glu Lys Val Ala Leu Gln Met Ile
            180                 185                 190

Glu Glu Leu Leu Asn Thr Asn
        195
```

<210> 29
<211> 2526
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (338)..(2275)
<223> G680

<400> 29

```
cagttatctt cttccttctt ctctctgttt tttaaattta tttttagaga atttttttg      60

ttttgcttcc gatttgatta tttccgggaa cgatgacttc tccggggagt tcccggtgag     120

atgataagtc agattgcata cttgtctcct ccatggctac tctcaagggt tttggctgcg     180

gtggattcgt ttggtttctc tagaatctaa agaggttatc acaacggctt tgcaatttga     240

aaactttcat gtttggggag atcaaagatg gtttcttttt tatactttac ttgttagaga     300

ggatttgaag cagcgaatag ctgcaaccgg tcctgtt atg gat act aat aca tct      355
                                        Met Asp Thr Asn Thr Ser
                                        1               5

gga gaa gaa tta tta gct aag gca aga aag cca tat aca ata aca aag      403
Gly Glu Glu Leu Leu Ala Lys Ala Arg Lys Pro Tyr Thr Ile Thr Lys
            10                  15                  20

cag cga gag cga tgg act gag gat gag cat gag agg ttt cta gaa gcc      451
Gln Arg Glu Arg Trp Thr Glu Asp Glu His Glu Arg Phe Leu Glu Ala
        25                  30                  35

ttg agg ctt tat gga aga gct tgg caa cga att gaa gaa cat att ggg      499
Leu Arg Leu Tyr Gly Arg Ala Trp Gln Arg Ile Glu Glu His Ile Gly
    40                  45                  50

aca aag act gct gtt cag atc aga agt cat gca caa aag ttc ttc aca      547
Thr Lys Thr Ala Val Gln Ile Arg Ser His Ala Gln Lys Phe Phe Thr
55                  60                  65                  70

aag ttg gag aaa gag gct gaa gtt aaa ggc atc cct gtt tgc caa gct      595
Lys Leu Glu Lys Glu Ala Glu Val Lys Gly Ile Pro Val Cys Gln Ala
                75                  80                  85

ttg gac ata gaa att ccg cct cct cgt cct aaa cga aaa ccc aat act      643
Leu Asp Ile Glu Ile Pro Pro Pro Arg Pro Lys Arg Lys Pro Asn Thr
            90                  95                  100

cct tat cct cga aaa cct ggg aac aac ggt aca tct tcc tct caa gta      691
Pro Tyr Pro Arg Lys Pro Gly Asn Asn Gly Thr Ser Ser Ser Gln Val
        105                 110                 115

tca tca gca aaa gat gca aaa ctt gtt tca tcg gcc tct tct tca cag      739
Ser Ser Ala Lys Asp Ala Lys Leu Val Ser Ser Ala Ser Ser Ser Gln
    120                 125                 130

ttg aat cag gcg ttc ttg gat ttg gaa aaa atg ccg ttc tct gag aaa      787
Leu Asn Gln Ala Phe Leu Asp Leu Glu Lys Met Pro Phe Ser Glu Lys
```

```
135                 140                 145                 150

aca tca act gga aaa gaa aat caa gat gag aat tgc tcg ggt gtt tct      835
Thr Ser Thr Gly Lys Glu Asn Gln Asp Glu Asn Cys Ser Gly Val Ser
                155                 160                 165

act gtg aac aag tat ccc tta cca acg aaa cag gta agt ggc gac att      883
Thr Val Asn Lys Tyr Pro Leu Pro Thr Lys Gln Val Ser Gly Asp Ile
            170                 175                 180

gaa aca agt aag acc tca act gtg gac aac gcg gtt caa gat gtt ccc      931
Glu Thr Ser Lys Thr Ser Thr Val Asp Asn Ala Val Gln Asp Val Pro
        185                 190                 195

aag aag aac aaa gac aaa gat ggt aac gat ggt act act gtg cac agc      979
Lys Lys Asn Lys Asp Lys Asp Gly Asn Asp Gly Thr Thr Val His Ser
    200                 205                 210

atg caa aac tac cct tgg cat ttc cac gca gat att gtg aac ggg aat     1027
Met Gln Asn Tyr Pro Trp His Phe His Ala Asp Ile Val Asn Gly Asn
215                 220                 225                 230

ata gca aaa tgc cct caa aat cat ccc tca ggt atg gta tct caa gac     1075
Ile Ala Lys Cys Pro Gln Asn His Pro Ser Gly Met Val Ser Gln Asp
                235                 240                 245

ttc atg ttt cat cct atg aga gaa gaa act cac ggg cac gca aat ctt     1123
Phe Met Phe His Pro Met Arg Glu Glu Thr His Gly His Ala Asn Leu
            250                 255                 260

caa gct aca aca gca tct gct act act aca gct tct cat caa gcg ttt     1171
Gln Ala Thr Thr Ala Ser Ala Thr Thr Thr Ala Ser His Gln Ala Phe
        265                 270                 275

cca gct tgt cat tca cag gat gat tac cgt tcg ttt ctc cag ata tca     1219
Pro Ala Cys His Ser Gln Asp Asp Tyr Arg Ser Phe Leu Gln Ile Ser
    280                 285                 290

tct act ttc tcc aat ctt att atg tca act ctc cta cag aat cct gca     1267
Ser Thr Phe Ser Asn Leu Ile Met Ser Thr Leu Leu Gln Asn Pro Ala
295                 300                 305                 310

gct cat gct gca gct aca ttc gct gct tcg gtc tgg cct tat gcg agt     1315
Ala His Ala Ala Ala Thr Phe Ala Ala Ser Val Trp Pro Tyr Ala Ser
                315                 320                 325

gtc ggg aat tct ggt gat tca tca acc cca atg agc tct tct cct cca     1363
Val Gly Asn Ser Gly Asp Ser Ser Thr Pro Met Ser Ser Ser Pro Pro
            330                 335                 340

agt ata act gcc att gcc gct gct aca gta gct gct gca act gct tgg     1411
Ser Ile Thr Ala Ile Ala Ala Ala Thr Val Ala Ala Ala Thr Ala Trp
        345                 350                 355

tgg gct tct cat gga ctt ctt cct gta tgc gct cca gct cca ata aca     1459
Trp Ala Ser His Gly Leu Leu Pro Val Cys Ala Pro Ala Pro Ile Thr
    360                 365                 370

tgt gtt cca ttc tca act gtt gca gtt cca act cca gca atg act gaa     1507
Cys Val Pro Phe Ser Thr Val Ala Val Pro Thr Pro Ala Met Thr Glu
375                 380                 385                 390

atg gat acc gtt gaa aat act caa ccg ttt gag aaa caa aac aca gct     1555
Met Asp Thr Val Glu Asn Thr Gln Pro Phe Glu Lys Gln Asn Thr Ala
                395                 400                 405

ctg caa gat caa acc ttg gct tcg aaa tct cca gct tca tca tct gat     1603
Leu Gln Asp Gln Thr Leu Ala Ser Lys Ser Pro Ala Ser Ser Ser Asp
            410                 415                 420

gat tca gat gag act gga gta acc aag cta aat gcc gac tca aaa acc     1651
Asp Ser Asp Glu Thr Gly Val Thr Lys Leu Asn Ala Asp Ser Lys Thr
        425                 430                 435

aat gat gat aaa att gag gag gtt gtt gtt act gcc gct gtg cat gac     1699
```

MBI-20 Sequence Listing.ST25

```
Asn Asp Asp Lys Ile Glu Glu Val Val Val Thr Ala Ala Val His Asp
    440                 445                 450

tca aac act gcc cag aag aaa aat ctt gtg gac cgc tca tcg tgt ggc     1747
Ser Asn Thr Ala Gln Lys Lys Asn Leu Val Asp Arg Ser Ser Cys Gly
455                 460                 465                 470

tca aat aca cct tca ggg agt gac gca gaa act gat gca tta gat aaa     1795
Ser Asn Thr Pro Ser Gly Ser Asp Ala Glu Thr Asp Ala Leu Asp Lys
                475                 480                 485

atg gag aaa gat aaa gag gat gtg aag gag aca gat gag aat cag cca     1843
Met Glu Lys Asp Lys Glu Asp Val Lys Glu Thr Asp Glu Asn Gln Pro
            490                 495                 500

gat gtt att gag tta aat aac cgt aag att aaa atg aga gac aac aac     1891
Asp Val Ile Glu Leu Asn Asn Arg Lys Ile Lys Met Arg Asp Asn Asn
        505                 510                 515

agc aac aac aat gca act act gat tcg tgg aag gaa gtc tcc gaa gag     1939
Ser Asn Asn Asn Ala Thr Thr Asp Ser Trp Lys Glu Val Ser Glu Glu
    520                 525                 530

ggt cgt ata gcg ttt cag gct ctc ttt gca aga gaa aga ttg cct caa     1987
Gly Arg Ile Ala Phe Gln Ala Leu Phe Ala Arg Glu Arg Leu Pro Gln
535                 540                 545                 550

agc ttt tcg cct cct caa gtg gca gag aat gtg aat aga aaa caa agt     2035
Ser Phe Ser Pro Pro Gln Val Ala Glu Asn Val Asn Arg Lys Gln Ser
                555                 560                 565

gac acg tca atg cca ttg gct cct aat ttc aaa agc cag gat tct tgt     2083
Asp Thr Ser Met Pro Leu Ala Pro Asn Phe Lys Ser Gln Asp Ser Cys
            570                 575                 580

gct gca gac caa gaa gga gta gta atg atc ggt gtt gga aca tgc aag     2131
Ala Ala Asp Gln Glu Gly Val Val Met Ile Gly Val Gly Thr Cys Lys
        585                 590                 595

agt ctt aaa acg aga cag aca gga ttt aag cca tac aag aga tgt tca     2179
Ser Leu Lys Thr Arg Gln Thr Gly Phe Lys Pro Tyr Lys Arg Cys Ser
    600                 605                 610

atg gaa gtg aaa gag agc caa gtt ggg aac ata aac aat caa agt gat     2227
Met Glu Val Lys Glu Ser Gln Val Gly Asn Ile Asn Asn Gln Ser Asp
615                 620                 625                 630

gaa aaa gtc tgc aaa agg ctt cga ttg gaa gga gaa gct tct aca tga     2275
Glu Lys Val Cys Lys Arg Leu Arg Leu Glu Gly Glu Ala Ser Thr
                635                 640                 645

cagacttgga ggtaaaaaaa aaacatccac atttttatca atatctttaa atctagtgtt   2335

agtagtttgc ttctccaatc tttatgaaag agacttttaa ttttccttcc gaacatttct   2395

ttggtcatgt caggttctgt accatattac cccatgtctt gtctcttgtc tctgtttgtg   2455

tatgctactt gtggtctata tgtcatctgc tactactgtt aattaaccat taagcaatgg   2515

atttgtcttt a                                                       2526
```

<210> 30
<211> 645
<212> PRT
<213> Arabidopsis thaliana

<400> 30

```
Met Asp Thr Asn Thr Ser Gly Glu Glu Leu Leu Ala Lys Ala Arg Lys
1               5                   10                  15

Pro Tyr Thr Ile Thr Lys Gln Arg Glu Arg Trp Thr Glu Asp Glu His
            20                  25                  30
```

Glu Arg Phe Leu Glu Ala Leu Arg Leu Tyr Gly Arg Ala Trp Gln Arg
35 40 45

Ile Glu Glu His Ile Gly Thr Lys Thr Ala Val Gln Ile Arg Ser His
50 55 60

Ala Gln Lys Phe Phe Thr Lys Leu Glu Lys Glu Ala Glu Val Lys Gly
65 70 75 80

Ile Pro Val Cys Gln Ala Leu Asp Ile Glu Ile Pro Pro Pro Arg Pro
85 90 95

Lys Arg Lys Pro Asn Thr Pro Tyr Pro Arg Lys Pro Gly Asn Asn Gly
100 105 110

Thr Ser Ser Ser Gln Val Ser Ser Ala Lys Asp Ala Lys Leu Val Ser
115 120 125

Ser Ala Ser Ser Ser Gln Leu Asn Gln Ala Phe Leu Asp Leu Glu Lys
130 135 140

Met Pro Phe Ser Glu Lys Thr Ser Thr Gly Lys Glu Asn Gln Asp Glu
145 150 155 160

Asn Cys Ser Gly Val Ser Thr Val Asn Lys Tyr Pro Leu Pro Thr Lys
165 170 175

Gln Val Ser Gly Asp Ile Glu Thr Ser Lys Thr Ser Thr Val Asp Asn
180 185 190

Ala Val Gln Asp Val Pro Lys Lys Asn Lys Asp Lys Asp Gly Asn Asp
195 200 205

Gly Thr Thr Val His Ser Met Gln Asn Tyr Pro Trp His Phe His Ala
210 215 220

Asp Ile Val Asn Gly Asn Ile Ala Lys Cys Pro Gln Asn His Pro Ser
225 230 235 240

Gly Met Val Ser Gln Asp Phe Met Phe His Pro Met Arg Glu Glu Thr
245 250 255

His Gly His Ala Asn Leu Gln Ala Thr Thr Ala Ser Ala Thr Thr Thr
260 265 270

Ala Ser His Gln Ala Phe Pro Ala Cys His Ser Gln Asp Asp Tyr Arg
275 280 285

Ser Phe Leu Gln Ile Ser Ser Thr Phe Ser Asn Leu Ile Met Ser Thr
290 295 300

Leu Leu Gln Asn Pro Ala Ala His Ala Ala Ala Thr Phe Ala Ala Ser
305 310 315 320

Val Trp Pro Tyr Ala Ser Val Gly Asn Ser Gly Asp Ser Ser Thr Pro

325 330 335

Met Ser Ser Ser Pro Pro Ser Ile Thr Ala Ile Ala Ala Ala Thr Val
340 345 350

Ala Ala Ala Thr Ala Trp Trp Ala Ser His Gly Leu Leu Pro Val Cys
355 360 365

Ala Pro Ala Pro Ile Thr Cys Val Pro Phe Ser Thr Val Ala Val Pro
370 375 380

Thr Pro Ala Met Thr Glu Met Asp Thr Val Glu Asn Thr Gln Pro Phe
385 390 395 400

Glu Lys Gln Asn Thr Ala Leu Gln Asp Gln Thr Leu Ala Ser Lys Ser
405 410 415

Pro Ala Ser Ser Ser Asp Asp Ser Asp Glu Thr Gly Val Thr Lys Leu
420 425 430

Asn Ala Asp Ser Lys Thr Asn Asp Asp Lys Ile Glu Glu Val Val Val
435 440 445

Thr Ala Ala Val His Asp Ser Asn Thr Ala Gln Lys Lys Asn Leu Val
450 455 460

Asp Arg Ser Ser Cys Gly Ser Asn Thr Pro Ser Gly Ser Asp Ala Glu
465 470 475 480

Thr Asp Ala Leu Asp Lys Met Glu Lys Asp Lys Glu Asp Val Lys Glu
485 490 495

Thr Asp Glu Asn Gln Pro Asp Val Ile Glu Leu Asn Asn Arg Lys Ile
500 505 510

Lys Met Arg Asp Asn Asn Ser Asn Asn Asn Ala Thr Thr Asp Ser Trp
515 520 525

Lys Glu Val Ser Glu Glu Gly Arg Ile Ala Phe Gln Ala Leu Phe Ala
530 535 540

Arg Glu Arg Leu Pro Gln Ser Phe Ser Pro Pro Gln Val Ala Glu Asn
545 550 555 560

Val Asn Arg Lys Gln Ser Asp Thr Ser Met Pro Leu Ala Pro Asn Phe
565 570 575

Lys Ser Gln Asp Ser Cys Ala Ala Asp Gln Glu Gly Val Val Met Ile
580 585 590

Gly Val Gly Thr Cys Lys Ser Leu Lys Thr Arg Gln Thr Gly Phe Lys
595 600 605

Pro Tyr Lys Arg Cys Ser Met Glu Val Lys Glu Ser Gln Val Gly Asn
610 615 620

```
Ile Asn Asn Gln Ser Asp Glu Lys Val Cys Lys Arg Leu Arg Leu Glu
625                 630                 635                 640

Gly Glu Ala Ser Thr
                645
```

<210> 31
<211> 1195
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (67)..(1041)
<223> G883

<400> 31

```
ctctctcgtc ttcgtcttct tcttcttcaa cgttcctctc caaaatcctc agaccaagaa      60

atcatc atg gcc gtc gat cta atg cgt ttc cct aag ata gat gat caa         108
       Met Ala Val Asp Leu Met Arg Phe Pro Lys Ile Asp Asp Gln
       1               5                   10

acg gct att cag gaa gct gca tcg caa ggt tta caa agt atg gaa cat       156
Thr Ala Ile Gln Glu Ala Ala Ser Gln Gly Leu Gln Ser Met Glu His
15                  20                  25                  30

ctg atc cgt gtc ctc tct aac cgt ccc gaa caa caa cac aac gtt gac       204
Leu Ile Arg Val Leu Ser Asn Arg Pro Glu Gln Gln His Asn Val Asp
                35                  40                  45

tgc tcc gag atc act gac ttc acc gtt tct aaa ttc aaa acc gtc att       252
Cys Ser Glu Ile Thr Asp Phe Thr Val Ser Lys Phe Lys Thr Val Ile
            50                  55                  60

tct ctc ctt aac cgt act ggt cac gct cgg ttc aga cgc gga ccg gtt       300
Ser Leu Leu Asn Arg Thr Gly His Ala Arg Phe Arg Arg Gly Pro Val
        65                  70                  75

cac tcc act tcc tct gcc gca tct cag aaa cta cag agt cag atc gtt       348
His Ser Thr Ser Ser Ala Ala Ser Gln Lys Leu Gln Ser Gln Ile Val
    80                  85                  90

aaa aat act caa cct gag gct ccg ata gtg aga aca act acg aat cac       396
Lys Asn Thr Gln Pro Glu Ala Pro Ile Val Arg Thr Thr Thr Asn His
95                  100                 105                 110

cct caa atc gtt cct cca ccg tct agt gta aca ctc gat ttc tct aaa       444
Pro Gln Ile Val Pro Pro Pro Ser Ser Val Thr Leu Asp Phe Ser Lys
                115                 120                 125

cca agc atc ttc ggc acc aaa gct aag agc gcc gag ctg gaa ttc tcc       492
Pro Ser Ile Phe Gly Thr Lys Ala Lys Ser Ala Glu Leu Glu Phe Ser
            130                 135                 140

aaa gaa aac ttc agt gtt tct tta aac tcc tca ttc atg tcg tcg gcg       540
Lys Glu Asn Phe Ser Val Ser Leu Asn Ser Ser Phe Met Ser Ser Ala
        145                 150                 155

ata acc gga gac ggc agc gtc tcc aat gga aaa atc ttc ctt gct tct       588
Ile Thr Gly Asp Gly Ser Val Ser Asn Gly Lys Ile Phe Leu Ala Ser
    160                 165                 170

gct ccg tcg cag cct gtt aac tct tcc gga aaa cca ccg ttg gct ggt       636
Ala Pro Ser Gln Pro Val Asn Ser Ser Gly Lys Pro Pro Leu Ala Gly
175                 180                 185                 190

cat cct tac aga aag aga tgt ctc gag cat gag cac tca gag agt ttc       684
His Pro Tyr Arg Lys Arg Cys Leu Glu His Glu His Ser Glu Ser Phe
                195                 200                 205

tcc gga aaa gtc tcc ggc tcc gcc tac gga aag tgc cat tgc aag aaa       732
```

MBI-20 Sequence Listing.ST25

```
Ser Gly Lys Val Ser Gly Ser Ala Tyr Gly Lys Cys His Cys Lys Lys
            210                 215                 220

agg aaa aat cgg atg aag aga acc gtg aga gta ccg gcg ata agt gca      780
Arg Lys Asn Arg Met Lys Arg Thr Val Arg Val Pro Ala Ile Ser Ala
        225                 230                 235

aag atc gcc gat att cca ccg gac gaa tat tcg tgg agg aag tac gga      828
Lys Ile Ala Asp Ile Pro Pro Asp Glu Tyr Ser Trp Arg Lys Tyr Gly
    240                 245                 250

caa aaa ccg atc aag ggc tca cca cac cca cgt ggt tac tac aag tgc      876
Gln Lys Pro Ile Lys Gly Ser Pro His Pro Arg Gly Tyr Tyr Lys Cys
255                 260                 265                 270

agt aca ttc aga gga tgt cca gcg agg aaa cac gtg gaa cga gca tta      924
Ser Thr Phe Arg Gly Cys Pro Ala Arg Lys His Val Glu Arg Ala Leu
                275                 280                 285

gat gat cca gcg atg ctt att gtg aca tac gaa gga gag cac cgt cat      972
Asp Asp Pro Ala Met Leu Ile Val Thr Tyr Glu Gly Glu His Arg His
            290                 295                 300

aac caa tcc gcg atg cag gag aat att tct tct tca ggc att aat gat     1020
Asn Gln Ser Ala Met Gln Glu Asn Ile Ser Ser Ser Gly Ile Asn Asp
        305                 310                 315

tta gtg ttt gcc tcg gct tga ctttttttg tactatttgt tttttgattt         1071
Leu Val Phe Ala Ser Ala
    320

tttgagtact ttagatggat tgaaatttgt aaatttttt attaagaaat caatttaaat    1131

agagaaaaat tagtggtggt gcaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa   1191

aaaa                                                                1195
```

<210> 32
<211> 324
<212> PRT
<213> Arabidopsis thaliana

<400> 32

```
Met Ala Val Asp Leu Met Arg Phe Pro Lys Ile Asp Asp Gln Thr Ala
1               5                   10                  15

Ile Gln Glu Ala Ala Ser Gln Gly Leu Gln Ser Met Glu His Leu Ile
            20                  25                  30

Arg Val Leu Ser Asn Arg Pro Glu Gln Gln His Asn Val Asp Cys Ser
        35                  40                  45

Glu Ile Thr Asp Phe Thr Val Ser Lys Phe Lys Thr Val Ile Ser Leu
    50                  55                  60

Leu Asn Arg Thr Gly His Ala Arg Phe Arg Arg Gly Pro Val His Ser
65                  70                  75                  80

Thr Ser Ser Ala Ala Ser Gln Lys Leu Gln Ser Gln Ile Val Lys Asn
                85                  90                  95

Thr Gln Pro Glu Ala Pro Ile Val Arg Thr Thr Thr Asn His Pro Gln
            100                 105                 110

Ile Val Pro Pro Pro Ser Ser Val Thr Leu Asp Phe Ser Lys Pro Ser
        115                 120                 125
```

Ile Phe Gly Thr Lys Ala Lys Ser Ala Glu Leu Glu Phe Ser Lys Glu
130 135 140

Asn Phe Ser Val Ser Leu Asn Ser Ser Phe Met Ser Ser Ala Ile Thr
145 150 155 160

Gly Asp Gly Ser Val Ser Asn Gly Lys Ile Phe Leu Ala Ser Ala Pro
165 170 175

Ser Gln Pro Val Asn Ser Ser Gly Lys Pro Pro Leu Ala Gly His Pro
180 185 190

Tyr Arg Lys Arg Cys Leu Glu His Glu His Ser Glu Ser Phe Ser Gly
195 200 205

Lys Val Ser Gly Ser Ala Tyr Gly Lys Cys His Cys Lys Lys Arg Lys
210 215 220

Asn Arg Met Lys Arg Thr Val Arg Val Pro Ala Ile Ser Ala Lys Ile
225 230 235 240

Ala Asp Ile Pro Pro Asp Glu Tyr Ser Trp Arg Lys Tyr Gly Gln Lys
245 250 255

Pro Ile Lys Gly Ser Pro His Pro Arg Gly Tyr Tyr Lys Cys Ser Thr
260 265 270

Phe Arg Gly Cys Pro Ala Arg Lys His Val Glu Arg Ala Leu Asp Asp
275 280 285

Pro Ala Met Leu Ile Val Thr Tyr Glu Gly Glu His Arg His Asn Gln
290 295 300

Ser Ala Met Gln Glu Asn Ile Ser Ser Ser Gly Ile Asn Asp Leu Val
305 310 315 320

Phe Ala Ser Ala

<210> 33
<211> 1902
<212> DNA
<213> Arabidopsis thalinana

<220>
<221> CDS
<222> (1)..(1902)
<223> G1855

<400> 33
atg gcg aaa gag aac agt ggt cat cat cac caa aca gaa gca aga aga 48
Met Ala Lys Glu Asn Ser Gly His His His Gln Thr Glu Ala Arg Arg
1 5 10 15

aag aaa cta act ttg att ctt ggt gta agt gga ctc tgc att ttg ttc 96
Lys Lys Leu Thr Leu Ile Leu Gly Val Ser Gly Leu Cys Ile Leu Phe
20 25 30

tat gtt tta ggt gca tgg caa gcc aat acc gtc cca tct tct atc tcg 144

MBI-20 Sequence Listing.ST25

```
Tyr Val Leu Gly Ala Trp Gln Ala Asn Thr Val Pro Ser Ser Ile Ser
        35                  40                  45

aag ctc gga tgc gag acg caa tca aac cct tct tcg tcc tct tcc tct      192
Lys Leu Gly Cys Glu Thr Gln Ser Asn Pro Ser Ser Ser Ser Ser Ser
    50                  55                  60

tcc tca tct tca gag tca gct gaa cta gat ttc aaa agc cat aat cag      240
Ser Ser Ser Ser Glu Ser Ala Glu Leu Asp Phe Lys Ser His Asn Gln
65                  70                  75                  80

att gag tta aag gaa aca aac caa acc att aag tac ttt gaa cca tgt      288
Ile Glu Leu Lys Glu Thr Asn Gln Thr Ile Lys Tyr Phe Glu Pro Cys
                85                  90                  95

gaa tta tct ctc agt gag tac act cct tgt gaa gac cga caa aga gga      336
Glu Leu Ser Leu Ser Glu Tyr Thr Pro Cys Glu Asp Arg Gln Arg Gly
            100                 105                 110

aga aga ttc gat agg aac atg atg aaa tat aga gaa aga cat tgt cct      384
Arg Arg Phe Asp Arg Asn Met Met Lys Tyr Arg Glu Arg His Cys Pro
        115                 120                 125

gta aaa gat gag ctt ctt tat tgt ttg att cct cct cca cca aac tac      432
Val Lys Asp Glu Leu Leu Tyr Cys Leu Ile Pro Pro Pro Pro Asn Tyr
    130                 135                 140

aag att cca ttt aaa tgg cca caa agt aga gac tat gct tgg tat gac      480
Lys Ile Pro Phe Lys Trp Pro Gln Ser Arg Asp Tyr Ala Trp Tyr Asp
145                 150                 155                 160

aat atc cct cac aag gaa ctt agt gtt gag aaa gca gtt caa aac tgg      528
Asn Ile Pro His Lys Glu Leu Ser Val Glu Lys Ala Val Gln Asn Trp
                165                 170                 175

att caa gtt gaa ggt gac cgc ttt aga ttc cct ggt ggt ggt act atg      576
Ile Gln Val Glu Gly Asp Arg Phe Arg Phe Pro Gly Gly Gly Thr Met
            180                 185                 190

ttt cct cgt gga gct gat gct tat atc gat gat att gct agg ctt att      624
Phe Pro Arg Gly Ala Asp Ala Tyr Ile Asp Asp Ile Ala Arg Leu Ile
        195                 200                 205

cct ctt act gat ggt gga atc aga aca gct att gac act gga tgt ggt      672
Pro Leu Thr Asp Gly Gly Ile Arg Thr Ala Ile Asp Thr Gly Cys Gly
    210                 215                 220

gtt gca agt ttt ggt gct tac ctc ttg aag aga gac att atg gct gtg      720
Val Ala Ser Phe Gly Ala Tyr Leu Leu Lys Arg Asp Ile Met Ala Val
225                 230                 235                 240

tct ttt gct cca aga gac act cat gaa gct cag gta cag ttt gct tta      768
Ser Phe Ala Pro Arg Asp Thr His Glu Ala Gln Val Gln Phe Ala Leu
                245                 250                 255

gaa cgc gga gtt cct gcg ata atc ggg att atg gga tca aga aga ctt      816
Glu Arg Gly Val Pro Ala Ile Ile Gly Ile Met Gly Ser Arg Arg Leu
            260                 265                 270

cct tat cca gct aga gct ttt gat ctt gct cat tgt tct cgt tgt ttg      864
Pro Tyr Pro Ala Arg Ala Phe Asp Leu Ala His Cys Ser Arg Cys Leu
        275                 280                 285

atc cct tgg ttt aaa aat gat ggt ttg tac ctt atg gag gtc gac cgg      912
Ile Pro Trp Phe Lys Asn Asp Gly Leu Tyr Leu Met Glu Val Asp Arg
    290                 295                 300

gtt tta aga ccg ggc ggt tac tgg atc ctc tcg gga cca ccg att aac      960
Val Leu Arg Pro Gly Gly Tyr Trp Ile Leu Ser Gly Pro Pro Ile Asn
305                 310                 315                 320

tgg aaa cag tac tgg aga ggg tgg gag aga aca gag gag gat ttg aag     1008
Trp Lys Gln Tyr Trp Arg Gly Trp Glu Arg Thr Glu Glu Asp Leu Lys
                325                 330                 335
```

```
aaa gag caa gat tca ata gaa gat gta gca aag agt ctt tgc tgg aag      1056
Lys Glu Gln Asp Ser Ile Glu Asp Val Ala Lys Ser Leu Cys Trp Lys
            340                 345                 350

aaa gta act gaa aaa ggt gac tta tca att tgg caa aag cct ctc aat      1104
Lys Val Thr Glu Lys Gly Asp Leu Ser Ile Trp Gln Lys Pro Leu Asn
        355                 360                 365

cac att gag tgt aaa aag ctc aaa caa aac aat aag tca cct ccg ata      1152
His Ile Glu Cys Lys Lys Leu Lys Gln Asn Asn Lys Ser Pro Pro Ile
    370                 375                 380

tgc agc tca gat aac gcg gat tcc gct tgg tac aaa gac ttg gaa act      1200
Cys Ser Ser Asp Asn Ala Asp Ser Ala Trp Tyr Lys Asp Leu Glu Thr
385                 390                 395                 400

tgt ata aca cca tta cca gaa aca aac aat cca gat gat tca gca ggc      1248
Cys Ile Thr Pro Leu Pro Glu Thr Asn Asn Pro Asp Asp Ser Ala Gly
                405                 410                 415

ggt gca ctc gag gat tgg cca gac cga gca ttc gcg gta cct cca aga      1296
Gly Ala Leu Glu Asp Trp Pro Asp Arg Ala Phe Ala Val Pro Pro Arg
            420                 425                 430

atc atc aga gga act ata cca gaa atg aac gcg gag aaa ttt aga gaa      1344
Ile Ile Arg Gly Thr Ile Pro Glu Met Asn Ala Glu Lys Phe Arg Glu
        435                 440                 445

gac aac gag gtt tgg aaa gag aga ata gca cat tac aag aag ata gtc      1392
Asp Asn Glu Val Trp Lys Glu Arg Ile Ala His Tyr Lys Lys Ile Val
    450                 455                 460

cct gag ctt tca cat gga aga ttc agg aac att atg gac atg aac gct      1440
Pro Glu Leu Ser His Gly Arg Phe Arg Asn Ile Met Asp Met Asn Ala
465                 470                 475                 480

ttt ctc ggc gga ttc gct gct tcc atg ctg aaa tat ccc tca tgg gtc      1488
Phe Leu Gly Gly Phe Ala Ala Ser Met Leu Lys Tyr Pro Ser Trp Val
                485                 490                 495

atg aac gtt gtc ccg gtc gat gca gag aaa caa acg tta ggt gtg atc      1536
Met Asn Val Val Pro Val Asp Ala Glu Lys Gln Thr Leu Gly Val Ile
            500                 505                 510

tac gaa cgt gga ttg ata ggg acg tat caa gat tgg tgt gaa gga ttc      1584
Tyr Glu Arg Gly Leu Ile Gly Thr Tyr Gln Asp Trp Cys Glu Gly Phe
        515                 520                 525

tca acg tat cca aga act tat gat atg att cat gca gga gga ttg ttc      1632
Ser Thr Tyr Pro Arg Thr Tyr Asp Met Ile His Ala Gly Gly Leu Phe
    530                 535                 540

agc tta tac gaa cat agg tgt gat ttg acg ttg ata ttg ttg gag atg      1680
Ser Leu Tyr Glu His Arg Cys Asp Leu Thr Leu Ile Leu Leu Glu Met
545                 550                 555                 560

gat cga att ttg aga cca gaa gga aca gtt gtg ttg aga gat aat gtg      1728
Asp Arg Ile Leu Arg Pro Glu Gly Thr Val Val Leu Arg Asp Asn Val
                565                 570                 575

gag acg ttg aat aag gta gag aag ata gtg aag gga atg aag tgg aag      1776
Glu Thr Leu Asn Lys Val Glu Lys Ile Val Lys Gly Met Lys Trp Lys
            580                 585                 590

agt caa att gtt gat cat gag aaa ggt cct ttt aat cct gag aag att      1824
Ser Gln Ile Val Asp His Glu Lys Gly Pro Phe Asn Pro Glu Lys Ile
        595                 600                 605

ctt gtt gct gtt aaa act tat tgg act ggt caa cct tct gac aag aac      1872
Leu Val Ala Val Lys Thr Tyr Trp Thr Gly Gln Pro Ser Asp Lys Asn
    610                 615                 620

aac aac aac aac aac aac aac aac aac tag                              1902
Asn Asn Asn Asn Asn Asn Asn Asn Asn
625                 630
```

MBI-20 Sequence Listing.ST25

<210> 34
<211> 633
<212> PRT
<213> Arabidopsis thalinana

<400> 34

Met Ala Lys Glu Asn Ser Gly His His His Gln Thr Glu Ala Arg Arg
1 5 10 15

Lys Lys Leu Thr Leu Ile Leu Gly Val Ser Gly Leu Cys Ile Leu Phe
20 25 30

Tyr Val Leu Gly Ala Trp Gln Ala Asn Thr Val Pro Ser Ser Ile Ser
35 40 45

Lys Leu Gly Cys Glu Thr Gln Ser Asn Pro Ser Ser Ser Ser Ser Ser
50 55 60

Ser Ser Ser Ser Glu Ser Ala Glu Leu Asp Phe Lys Ser His Asn Gln
65 70 75 80

Ile Glu Leu Lys Glu Thr Asn Gln Thr Ile Lys Tyr Phe Glu Pro Cys
85 90 95

Glu Leu Ser Leu Ser Glu Tyr Thr Pro Cys Glu Asp Arg Gln Arg Gly
100 105 110

Arg Arg Phe Asp Arg Asn Met Met Lys Tyr Arg Glu Arg His Cys Pro
115 120 125

Val Lys Asp Glu Leu Leu Tyr Cys Leu Ile Pro Pro Pro Pro Asn Tyr
130 135 140

Lys Ile Pro Phe Lys Trp Pro Gln Ser Arg Asp Tyr Ala Trp Tyr Asp
145 150 155 160

Asn Ile Pro His Lys Glu Leu Ser Val Glu Lys Ala Val Gln Asn Trp
165 170 175

Ile Gln Val Glu Gly Asp Arg Phe Arg Phe Pro Gly Gly Gly Thr Met
180 185 190

Phe Pro Arg Gly Ala Asp Ala Tyr Ile Asp Asp Ile Ala Arg Leu Ile
195 200 205

Pro Leu Thr Asp Gly Gly Ile Arg Thr Ala Ile Asp Thr Gly Cys Gly
210 215 220

Val Ala Ser Phe Gly Ala Tyr Leu Leu Lys Arg Asp Ile Met Ala Val
225 230 235 240

Ser Phe Ala Pro Arg Asp Thr His Glu Ala Gln Val Gln Phe Ala Leu
245 250 255

Glu Arg Gly Val Pro Ala Ile Ile Gly Ile Met Gly Ser Arg Arg Leu
260 265 270

```
Pro Tyr Pro Ala Arg Ala Phe Asp Leu Ala His Cys Ser Arg Cys Leu
        275                 280                 285

Ile Pro Trp Phe Lys Asn Asp Gly Leu Tyr Leu Met Glu Val Asp Arg
    290                 295                 300

Val Leu Arg Pro Gly Gly Tyr Trp Ile Leu Ser Gly Pro Pro Ile Asn
305                 310                 315                 320

Trp Lys Gln Tyr Trp Arg Gly Trp Glu Arg Thr Glu Glu Asp Leu Lys
                325                 330                 335

Lys Glu Gln Asp Ser Ile Glu Asp Val Ala Lys Ser Leu Cys Trp Lys
            340                 345                 350

Lys Val Thr Glu Lys Gly Asp Leu Ser Ile Trp Gln Lys Pro Leu Asn
        355                 360                 365

His Ile Glu Cys Lys Lys Leu Lys Gln Asn Asn Lys Ser Pro Pro Ile
    370                 375                 380

Cys Ser Ser Asp Asn Ala Asp Ser Ala Trp Tyr Lys Asp Leu Glu Thr
385                 390                 395                 400

Cys Ile Thr Pro Leu Pro Glu Thr Asn Asn Pro Asp Asp Ser Ala Gly
                405                 410                 415

Gly Ala Leu Glu Asp Trp Pro Asp Arg Ala Phe Ala Val Pro Pro Arg
            420                 425                 430

Ile Ile Arg Gly Thr Ile Pro Glu Met Asn Ala Glu Lys Phe Arg Glu
        435                 440                 445

Asp Asn Glu Val Trp Lys Glu Arg Ile Ala His Tyr Lys Lys Ile Val
    450                 455                 460

Pro Glu Leu Ser His Gly Arg Phe Arg Asn Ile Met Asp Met Asn Ala
465                 470                 475                 480

Phe Leu Gly Gly Phe Ala Ala Ser Met Leu Lys Tyr Pro Ser Trp Val
                485                 490                 495

Met Asn Val Val Pro Val Asp Ala Glu Lys Gln Thr Leu Gly Val Ile
            500                 505                 510

Tyr Glu Arg Gly Leu Ile Gly Thr Tyr Gln Asp Trp Cys Glu Gly Phe
        515                 520                 525

Ser Thr Tyr Pro Arg Thr Tyr Asp Met Ile His Ala Gly Gly Leu Phe
    530                 535                 540

Ser Leu Tyr Glu His Arg Cys Asp Leu Thr Leu Ile Leu Leu Glu Met
545                 550                 555                 560

Asp Arg Ile Leu Arg Pro Glu Gly Thr Val Val Leu Arg Asp Asn Val
```

565 570 575

Glu Thr Leu Asn Lys Val Glu Lys Ile Val Lys Gly Met Lys Trp Lys
580 585 590

Ser Gln Ile Val Asp His Glu Lys Gly Pro Phe Asn Pro Glu Lys Ile
595 600 605

Leu Val Ala Val Lys Thr Tyr Trp Thr Gly Gln Pro Ser Asp Lys Asn
610 615 620

Asn Asn Asn Asn Asn Asn Asn Asn Asn
625 630

<210> 35
<211> 2324
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (209)..(2020)
<223> G1190

<400> 35
tcctgtccca aaaccaaaag gcttgagagt gtgtctttag agagagatct tctctctttt 60

atcttacgac tctcacttct tatctcaaat ctacttcaac tctatttcca gtctccacat 120

tttcccacaa atttcaactc ttgttctctt catccaaagt aaaaaacaaa tcgttgcaag 180

tgaggtttgg ttttggtgtt atagaatt atg aag agc ggg aag caa tct tcg 232
Met Lys Ser Gly Lys Gln Ser Ser
1 5

caa cct gaa aag ggt act tcc agg atc ttg tca ctg act gtc ctg ttt 280
Gln Pro Glu Lys Gly Thr Ser Arg Ile Leu Ser Leu Thr Val Leu Phe
10 15 20

atc gca ttt tgc ggt ttc tcc ttc tac ctc ggt ggt ata ttt tgc tct 328
Ile Ala Phe Cys Gly Phe Ser Phe Tyr Leu Gly Gly Ile Phe Cys Ser
25 30 35 40

gag aga gac aag att gta gcc aag gat gtc aca agg acg act aca aag 376
Glu Arg Asp Lys Ile Val Ala Lys Asp Val Thr Arg Thr Thr Thr Lys
45 50 55

gct gta gct tcc cct aaa gaa cct aca gct act cct att caa atc aaa 424
Ala Val Ala Ser Pro Lys Glu Pro Thr Ala Thr Pro Ile Gln Ile Lys
60 65 70

tcc gtt tct ttc ccg gag tgc ggg tca gag ttc caa gat tac acc ccg 472
Ser Val Ser Phe Pro Glu Cys Gly Ser Glu Phe Gln Asp Tyr Thr Pro
75 80 85

tgc acc gat cca aag agg tgg aag aag tat ggt gtc cat cgc tta agt 520
Cys Thr Asp Pro Lys Arg Trp Lys Lys Tyr Gly Val His Arg Leu Ser
90 95 100

ttc ttg gag cgt cat tgt cct ccg gta tat gaa aag aat gag tgt ttg 568
Phe Leu Glu Arg His Cys Pro Pro Val Tyr Glu Lys Asn Glu Cys Leu
105 110 115 120

att cca cca cca gac ggg tat aaa ccg cct ata aga tgg ccc aag agc 616
Ile Pro Pro Pro Asp Gly Tyr Lys Pro Pro Ile Arg Trp Pro Lys Ser
125 130 135

cga gaa cag tgt tgg tac agg aac gtg cct tat gat tgg atc aat aag 664
Arg Glu Gln Cys Trp Tyr Arg Asn Val Pro Tyr Asp Trp Ile Asn Lys

```
                140                 145                 150

caa aag tct aac cag cat tgg ctt aag aaa gaa gga gat aag ttc cat      712
Gln Lys Ser Asn Gln His Trp Leu Lys Lys Glu Gly Asp Lys Phe His
        155                 160                 165

ttc cct ggt ggt ggt acc atg ttc cct cgt gga gtt agt cac tat gtt      760
Phe Pro Gly Gly Gly Thr Met Phe Pro Arg Gly Val Ser His Tyr Val
    170                 175                 180

gat ttg atg caa gat ctg att cct gaa atg aaa gac gga aca gtc agg      808
Asp Leu Met Gln Asp Leu Ile Pro Glu Met Lys Asp Gly Thr Val Arg
185                 190                 195                 200

acc gcc att gat act ggc tgt ggg gtt gcg agc tgg gga ggc gat ctt      856
Thr Ala Ile Asp Thr Gly Cys Gly Val Ala Ser Trp Gly Gly Asp Leu
                205                 210                 215

ttg gac cgt ggg ata cta tca ctc tct ctt gct cca aga gat aac cat      904
Leu Asp Arg Gly Ile Leu Ser Leu Ser Leu Ala Pro Arg Asp Asn His
            220                 225                 230

gaa gct cag gtt caa ttt gct ctt gaa cgt gga att cct gcg att ctc      952
Glu Ala Gln Val Gln Phe Ala Leu Glu Arg Gly Ile Pro Ala Ile Leu
        235                 240                 245

ggg atc atc tct acg caa cgt ctc cct ttt cct tca aat gca ttt gat     1000
Gly Ile Ile Ser Thr Gln Arg Leu Pro Phe Pro Ser Asn Ala Phe Asp
    250                 255                 260

atg gct cat tgt tca aga tgt ctt att ccc tgg aca gaa ttt ggt gga     1048
Met Ala His Cys Ser Arg Cys Leu Ile Pro Trp Thr Glu Phe Gly Gly
265                 270                 275                 280

atc tat tta ctt gag att cac cgt ata gtt cga cct gga ggt ttt tgg     1096
Ile Tyr Leu Leu Glu Ile His Arg Ile Val Arg Pro Gly Gly Phe Trp
                285                 290                 295

gtt ctt tct ggt cca cct gtg aac tat aat aga cga tgg cgt gga tgg     1144
Val Leu Ser Gly Pro Pro Val Asn Tyr Asn Arg Arg Trp Arg Gly Trp
            300                 305                 310

aac aca acc atg gaa gat cag aaa tct gac tac aac aag ctt cag tca     1192
Asn Thr Thr Met Glu Asp Gln Lys Ser Asp Tyr Asn Lys Leu Gln Ser
        315                 320                 325

ctt cta acc tcc atg tgt ttc aaa aag tac gct caa aaa gat gac ata     1240
Leu Leu Thr Ser Met Cys Phe Lys Lys Tyr Ala Gln Lys Asp Asp Ile
    330                 335                 340

gcc gtg tgg cag aaa ctc tca gac aaa tct tgc tat gac aaa atc gct     1288
Ala Val Trp Gln Lys Leu Ser Asp Lys Ser Cys Tyr Asp Lys Ile Ala
345                 350                 355                 360

aag aac atg gaa gct tac cct ccc aaa tgt gac gac agt ata gaa cct     1336
Lys Asn Met Glu Ala Tyr Pro Pro Lys Cys Asp Asp Ser Ile Glu Pro
                365                 370                 375

gat tct gct tgg tac act cca ctc cgt cct tgc gtg gtt gcc ccg aca     1384
Asp Ser Ala Trp Tyr Thr Pro Leu Arg Pro Cys Val Val Ala Pro Thr
            380                 385                 390

cct aaa gtc aag aag tct ggt ctc gga tca atc cca aaa tgg ccc gag     1432
Pro Lys Val Lys Lys Ser Gly Leu Gly Ser Ile Pro Lys Trp Pro Glu
        395                 400                 405

agg tta cat gtc gcg ccc gag aga atc ggt gat gtt cac gga ggg agt     1480
Arg Leu His Val Ala Pro Glu Arg Ile Gly Asp Val His Gly Gly Ser
    410                 415                 420

gcg aac agt ttg aaa cac gat gat ggt aaa tgg aag aac aga gtt aag     1528
Ala Asn Ser Leu Lys His Asp Asp Gly Lys Trp Lys Asn Arg Val Lys
425                 430                 435                 440

cat tac aag aaa gtt tta cca gct ctt ggg aca gac aag ata aga aat     1576
```

```
His Tyr Lys Lys Val Leu Pro Ala Leu Gly Thr Asp Lys Ile Arg Asn
                445                 450                 455

gtt atg gat atg aac act gtt tat gga ggt ttc tct gcg gcc ctc att     1624
Val Met Asp Met Asn Thr Val Tyr Gly Gly Phe Ser Ala Ala Leu Ile
            460                 465                 470

gag gat ccc att tgg gtc atg aac gtt gta tca tcg tac agc gca aat     1672
Glu Asp Pro Ile Trp Val Met Asn Val Val Ser Ser Tyr Ser Ala Asn
        475                 480                 485

tcg ctt cct gtt gtc ttt gat cgc ggt ctc atc ggg act tac cac gac     1720
Ser Leu Pro Val Val Phe Asp Arg Gly Leu Ile Gly Thr Tyr His Asp
    490                 495                 500

tgg tgc gaa gct ttc tca acg tat cca aga aca tat gat ctt ctt cac     1768
Trp Cys Glu Ala Phe Ser Thr Tyr Pro Arg Thr Tyr Asp Leu Leu His
505                 510                 515                 520

ctc gac agt ctt ttt acc ttg gag agt cac agg tgt gag atg aag tac     1816
Leu Asp Ser Leu Phe Thr Leu Glu Ser His Arg Cys Glu Met Lys Tyr
                525                 530                 535

att ttg cta gag atg gac agg atc ttg cgg ccg agt gga tat gtt ata     1864
Ile Leu Leu Glu Met Asp Arg Ile Leu Arg Pro Ser Gly Tyr Val Ile
            540                 545                 550

atc cga gaa tcg agt tat ttc atg gac gca atc aca acg tta gcg aaa     1912
Ile Arg Glu Ser Ser Tyr Phe Met Asp Ala Ile Thr Thr Leu Ala Lys
        555                 560                 565

ggg ata agg tgg agt tgc cgg aga gag gag act gag tat gca gtc aaa     1960
Gly Ile Arg Trp Ser Cys Arg Arg Glu Glu Thr Glu Tyr Ala Val Lys
    570                 575                 580

agt gag aag att ctg gtt tgc cag aaa aag cta tgg ttt tcg tca aac     2008
Ser Glu Lys Ile Leu Val Cys Gln Lys Lys Leu Trp Phe Ser Ser Asn
585                 590                 595                 600

caa acc tct tga tgagaccacc tgtatcatag tgtttatcat ctcctgtgat         2060
Gln Thr Ser

gcacactaca gagagaagga tctagtcctt tgagtccaag atatagctct ataaacaatc   2120

tccttttttt gttctcttta atttcttggg tatttcacgg tatagattga tattatatat   2180

tttttaatta tatttttaat atatagatat attagtatgt ggtttaaaca ctattattat   2240

caaggtctta aagatttgct ttgcaagagt taaaaaatgt tggagtaagg acctcttgat   2300

taataaattg actgacgcag caaa                                         2324
```

<210> 36
<211> 603
<212> PRT
<213> Arabidopsis thaliana

<400> 36

```
Met Lys Ser Gly Lys Gln Ser Ser Gln Pro Glu Lys Gly Thr Ser Arg
1               5                   10                  15

Ile Leu Ser Leu Thr Val Leu Phe Ile Ala Phe Cys Gly Phe Ser Phe
            20                  25                  30

Tyr Leu Gly Gly Ile Phe Cys Ser Glu Arg Asp Lys Ile Val Ala Lys
        35                  40                  45

Asp Val Thr Arg Thr Thr Thr Lys Ala Val Ala Ser Pro Lys Glu Pro
    50                  55                  60
```

Thr Ala Thr Pro Ile Gln Ile Lys Ser Val Ser Phe Pro Glu Cys Gly
65 70 75 80

Ser Glu Phe Gln Asp Tyr Thr Pro Cys Thr Asp Pro Lys Arg Trp Lys
85 90 95

Lys Tyr Gly Val His Arg Leu Ser Phe Leu Glu Arg His Cys Pro Pro
100 105 110

Val Tyr Glu Lys Asn Glu Cys Leu Ile Pro Pro Pro Asp Gly Tyr Lys
115 120 125

Pro Pro Ile Arg Trp Pro Lys Ser Arg Glu Gln Cys Trp Tyr Arg Asn
130 135 140

Val Pro Tyr Asp Trp Ile Asn Lys Gln Lys Ser Asn Gln His Trp Leu
145 150 155 160

Lys Lys Glu Gly Asp Lys Phe His Phe Pro Gly Gly Gly Thr Met Phe
165 170 175

Pro Arg Gly Val Ser His Tyr Val Asp Leu Met Gln Asp Leu Ile Pro
180 185 190

Glu Met Lys Asp Gly Thr Val Arg Thr Ala Ile Asp Thr Gly Cys Gly
195 200 205

Val Ala Ser Trp Gly Gly Asp Leu Leu Asp Arg Gly Ile Leu Ser Leu
210 215 220

Ser Leu Ala Pro Arg Asp Asn His Glu Ala Gln Val Gln Phe Ala Leu
225 230 235 240

Glu Arg Gly Ile Pro Ala Ile Leu Gly Ile Ile Ser Thr Gln Arg Leu
245 250 255

Pro Phe Pro Ser Asn Ala Phe Asp Met Ala His Cys Ser Arg Cys Leu
260 265 270

Ile Pro Trp Thr Glu Phe Gly Gly Ile Tyr Leu Leu Glu Ile His Arg
275 280 285

Ile Val Arg Pro Gly Gly Phe Trp Val Leu Ser Gly Pro Pro Val Asn
290 295 300

Tyr Asn Arg Arg Trp Arg Gly Trp Asn Thr Thr Met Glu Asp Gln Lys
305 310 315 320

Ser Asp Tyr Asn Lys Leu Gln Ser Leu Leu Thr Ser Met Cys Phe Lys
325 330 335

Lys Tyr Ala Gln Lys Asp Asp Ile Ala Val Trp Gln Lys Leu Ser Asp
340 345 350

Lys Ser Cys Tyr Asp Lys Ile Ala Lys Asn Met Glu Ala Tyr Pro Pro

355 360 365

Lys Cys Asp Asp Ser Ile Glu Pro Asp Ser Ala Trp Tyr Thr Pro Leu
370 375 380

Arg Pro Cys Val Val Ala Pro Thr Pro Lys Val Lys Lys Ser Gly Leu
385 390 395 400

Gly Ser Ile Pro Lys Trp Pro Glu Arg Leu His Val Ala Pro Glu Arg
405 410 415

Ile Gly Asp Val His Gly Gly Ser Ala Asn Ser Leu Lys His Asp Asp
420 425 430

Gly Lys Trp Lys Asn Arg Val Lys His Tyr Lys Lys Val Leu Pro Ala
435 440 445

Leu Gly Thr Asp Lys Ile Arg Asn Val Met Asp Met Asn Thr Val Tyr
450 455 460

Gly Gly Phe Ser Ala Ala Leu Ile Glu Asp Pro Ile Trp Val Met Asn
465 470 475 480

Val Val Ser Ser Tyr Ser Ala Asn Ser Leu Pro Val Val Phe Asp Arg
485 490 495

Gly Leu Ile Gly Thr Tyr His Asp Trp Cys Glu Ala Phe Ser Thr Tyr
500 505 510

Pro Arg Thr Tyr Asp Leu Leu His Leu Asp Ser Leu Phe Thr Leu Glu
515 520 525

Ser His Arg Cys Glu Met Lys Tyr Ile Leu Leu Glu Met Asp Arg Ile
530 535 540

Leu Arg Pro Ser Gly Tyr Val Ile Ile Arg Glu Ser Ser Tyr Phe Met
545 550 555 560

Asp Ala Ile Thr Thr Leu Ala Lys Gly Ile Arg Trp Ser Cys Arg Arg
565 570 575

Glu Glu Thr Glu Tyr Ala Val Lys Ser Glu Lys Ile Leu Val Cys Gln
580 585 590

Lys Lys Leu Trp Phe Ser Ser Asn Gln Thr Ser
595 600

<210> 37
<211> 1951
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (196)..(1794)
<223> G308

<400> 37

MBI-20 Sequence Listing.ST25

```
agtaatttag ttttttttt ttttttttac aatttatttt gttattagaa gtggtagtgg      60

agtgaaaaaa caaatcctaa gcagtcctaa ccgatccccg aagctaaaga ttcttcacct     120

tcccaaataa agcaaaacct agatccgaca ttgaaggaaa aacctttag atccatctct      180

gaaaaaaacc caacc atg aag aga gat cat cat cat cat cat caa gat aag       231
                Met Lys Arg Asp His His His His His Gln Asp Lys
                1               5                   10

aag act atg atg atg aat gaa gaa gac gac ggt aac ggc atg gat gag       279
Lys Thr Met Met Met Asn Glu Glu Asp Asp Gly Asn Gly Met Asp Glu
        15                  20                  25

ctt cta gct gtt ctt ggt tac aag gtt agg tca tcg gaa atg gct gat       327
Leu Leu Ala Val Leu Gly Tyr Lys Val Arg Ser Ser Glu Met Ala Asp
    30                  35                  40

gtt gct cag aaa ctc gag cag ctt gaa gtt atg atg tct aat gtt caa       375
Val Ala Gln Lys Leu Glu Gln Leu Glu Val Met Met Ser Asn Val Gln
45                  50                  55                  60

gaa gac gat ctt tct caa ctc gct act gag act gtt cac tat aat ccg       423
Glu Asp Asp Leu Ser Gln Leu Ala Thr Glu Thr Val His Tyr Asn Pro
                65                  70                  75

gcg gag ctt tac acg tgg ctt gat tct atg ctc acc gac ctt aat cct       471
Ala Glu Leu Tyr Thr Trp Leu Asp Ser Met Leu Thr Asp Leu Asn Pro
            80                  85                  90

ccg tcg tct aac gcc gag tac gat ctt aaa gct att ccc ggt gac gcg       519
Pro Ser Ser Asn Ala Glu Tyr Asp Leu Lys Ala Ile Pro Gly Asp Ala
        95                  100                 105

att ctc aat cag ttc gct atc gat tcg gct tct tcg tct aac caa ggc       567
Ile Leu Asn Gln Phe Ala Ile Asp Ser Ala Ser Ser Ser Asn Gln Gly
    110                 115                 120

ggc gga gga gat acg tat act aca aac aag cgg ttg aaa tgc tca aac       615
Gly Gly Gly Asp Thr Tyr Thr Thr Asn Lys Arg Leu Lys Cys Ser Asn
125                 130                 135                 140

ggc gtc gtg gaa acc acc aca gcg acg gct gag tca act cgg cat gtt       663
Gly Val Val Glu Thr Thr Thr Ala Thr Ala Glu Ser Thr Arg His Val
                145                 150                 155

gtc ctg gtt gac tcg cag gag aac ggt gtg cgt ctc gtt cac gcg ctt       711
Val Leu Val Asp Ser Gln Glu Asn Gly Val Arg Leu Val His Ala Leu
            160                 165                 170

ttg gct tgc gct gaa gct gtt cag aag gag aat ctg act gtg gcg gaa       759
Leu Ala Cys Ala Glu Ala Val Gln Lys Glu Asn Leu Thr Val Ala Glu
        175                 180                 185

gct ctg gtg aag caa atc gga ttc tta gct gtt tct caa atc gga gct       807
Ala Leu Val Lys Gln Ile Gly Phe Leu Ala Val Ser Gln Ile Gly Ala
    190                 195                 200

atg aga caa gtc gct act tac ttc gcc gaa gct ctc gcg cgg cgg att       855
Met Arg Gln Val Ala Thr Tyr Phe Ala Glu Ala Leu Ala Arg Arg Ile
205                 210                 215                 220

tac cgt ctc tct ccg tcg cag agt cca atc gac cac tct ctc tcc gat       903
Tyr Arg Leu Ser Pro Ser Gln Ser Pro Ile Asp His Ser Leu Ser Asp
                225                 230                 235

act ctt cag atg cac ttc tac gag act tgt cct tat ctc aag ttc gct       951
Thr Leu Gln Met His Phe Tyr Glu Thr Cys Pro Tyr Leu Lys Phe Ala
            240                 245                 250

cac ttc acg gcg aat caa gcg att ctc gaa gct ttt caa ggg aag aaa       999
His Phe Thr Ala Asn Gln Ala Ile Leu Glu Ala Phe Gln Gly Lys Lys
        255                 260                 265

aga gtt cat gtc att gat ttc tct atg agt caa ggt ctt caa tgg ccg      1047
```

MBI-20 Sequence Listing.ST25

```
Arg Val His Val Ile Asp Phe Ser Met Ser Gln Gly Leu Gln Trp Pro
    270                 275                 280

gcg ctt atg cag gct ctt gcg ctt cga cct ggt ggt cct cct gtt ttc     1095
Ala Leu Met Gln Ala Leu Ala Leu Arg Pro Gly Gly Pro Pro Val Phe
285                 290                 295                 300

cgg tta acc gga att ggt cca ccg gca ccg gat aat ttc gat tat ctt     1143
Arg Leu Thr Gly Ile Gly Pro Pro Ala Pro Asp Asn Phe Asp Tyr Leu
                305                 310                 315

cat gaa gtt ggg tgt aag ctg gct cat tta gct gag gcg att cac gtt     1191
His Glu Val Gly Cys Lys Leu Ala His Leu Ala Glu Ala Ile His Val
            320                 325                 330

gag ttt gag tac aga gga ttt gtg gct aac act tta gct gat ctt gat     1239
Glu Phe Glu Tyr Arg Gly Phe Val Ala Asn Thr Leu Ala Asp Leu Asp
        335                 340                 345

gct tcg atg ctt gag ctt aga cca agt gag att gaa tct gtt gcg gtt     1287
Ala Ser Met Leu Glu Leu Arg Pro Ser Glu Ile Glu Ser Val Ala Val
    350                 355                 360

aac tct gtt ttc gag ctt cac aag ctc ttg gga cga cct ggt gcg atc     1335
Asn Ser Val Phe Glu Leu His Lys Leu Leu Gly Arg Pro Gly Ala Ile
365                 370                 375                 380

gat aag gtt ctt ggt gtg gtg aat cag att aaa ccg gag att ttc act     1383
Asp Lys Val Leu Gly Val Val Asn Gln Ile Lys Pro Glu Ile Phe Thr
                385                 390                 395

gtg gtt gag cag gaa tcg aac cat aat agt ccg att ttc tta gat cgg     1431
Val Val Glu Gln Glu Ser Asn His Asn Ser Pro Ile Phe Leu Asp Arg
            400                 405                 410

ttt act gag tcg ttg cat tat tac tcg acg ttg ttt gac tcg ttg gaa     1479
Phe Thr Glu Ser Leu His Tyr Tyr Ser Thr Leu Phe Asp Ser Leu Glu
        415                 420                 425

ggt gta ccg agt ggt caa gac aag gtc atg tcg gag gtt tac ttg ggt     1527
Gly Val Pro Ser Gly Gln Asp Lys Val Met Ser Glu Val Tyr Leu Gly
    430                 435                 440

aaa cag atc tgc aac gtt gtg gct tgt gat gga cct gac cga gtt gag     1575
Lys Gln Ile Cys Asn Val Val Ala Cys Asp Gly Pro Asp Arg Val Glu
445                 450                 455                 460

cgt cat gaa acg ttg agt cag tgg agg aac cgg ttc ggg tct gct ggg     1623
Arg His Glu Thr Leu Ser Gln Trp Arg Asn Arg Phe Gly Ser Ala Gly
                465                 470                 475

ttt gcg gct gca cat att ggt tcg aat gcg ttt aag caa gcg agt atg     1671
Phe Ala Ala Ala His Ile Gly Ser Asn Ala Phe Lys Gln Ala Ser Met
            480                 485                 490

ctt ttg gct ctg ttc aac ggc ggt gag ggt tat cgg gtg gag gag agt     1719
Leu Leu Ala Leu Phe Asn Gly Gly Glu Gly Tyr Arg Val Glu Glu Ser
        495                 500                 505

gac ggc tgt ctc atg ttg ggt tgg cac aca cga ccg ctc ata gcc acc     1767
Asp Gly Cys Leu Met Leu Gly Trp His Thr Arg Pro Leu Ile Ala Thr
    510                 515                 520

tcg gct tgg aaa ctc tcc acc aat tag atggtggctc aatgaattga           1814
Ser Ala Trp Lys Leu Ser Thr Asn
525                 530

tctgttgaac cggttatgat gatagatttc cgaccgaagc caaactaaat cctactgttt   1874

ttccctttgt cacttgttaa gatcttatct ttcattatat taggtaattg aaaaatttta   1934

atctcgccta attact                                                  1951
```

<210> 38

MBI-20 Sequence Listing.ST25

<211> 532
<212> PRT
<213> Arabidopsis thaliana

<400> 38

```
Met Lys Arg Asp His His His His His Gln Asp Lys Lys Thr Met Met
1               5                  10                  15

Met Asn Glu Glu Asp Asp Gly Asn Gly Met Asp Glu Leu Leu Ala Val
            20                  25                  30

Leu Gly Tyr Lys Val Arg Ser Ser Glu Met Ala Asp Val Ala Gln Lys
        35                  40                  45

Leu Glu Gln Leu Glu Val Met Met Ser Asn Val Gln Glu Asp Asp Leu
    50                  55                  60

Ser Gln Leu Ala Thr Glu Thr Val His Tyr Asn Pro Ala Glu Leu Tyr
65                  70                  75                  80

Thr Trp Leu Asp Ser Met Leu Thr Asp Leu Asn Pro Pro Ser Ser Asn
                85                  90                  95

Ala Glu Tyr Asp Leu Lys Ala Ile Pro Gly Asp Ala Ile Leu Asn Gln
            100                 105                 110

Phe Ala Ile Asp Ser Ala Ser Ser Ser Asn Gln Gly Gly Gly Gly Asp
        115                 120                 125

Thr Tyr Thr Thr Asn Lys Arg Leu Lys Cys Ser Asn Gly Val Val Glu
    130                 135                 140

Thr Thr Thr Ala Thr Ala Glu Ser Thr Arg His Val Val Leu Val Asp
145                 150                 155                 160

Ser Gln Glu Asn Gly Val Arg Leu Val His Ala Leu Leu Ala Cys Ala
                165                 170                 175

Glu Ala Val Gln Lys Glu Asn Leu Thr Val Ala Glu Ala Leu Val Lys
            180                 185                 190

Gln Ile Gly Phe Leu Ala Val Ser Gln Ile Gly Ala Met Arg Gln Val
        195                 200                 205

Ala Thr Tyr Phe Ala Glu Ala Leu Ala Arg Arg Ile Tyr Arg Leu Ser
    210                 215                 220

Pro Ser Gln Ser Pro Ile Asp His Ser Leu Ser Asp Thr Leu Gln Met
225                 230                 235                 240

His Phe Tyr Glu Thr Cys Pro Tyr Leu Lys Phe Ala His Phe Thr Ala
                245                 250                 255

Asn Gln Ala Ile Leu Glu Ala Phe Gln Gly Lys Lys Arg Val His Val
            260                 265                 270

Ile Asp Phe Ser Met Ser Gln Gly Leu Gln Trp Pro Ala Leu Met Gln
```

275 280 285

Ala Leu Ala Leu Arg Pro Gly Gly Pro Pro Val Phe Arg Leu Thr Gly
290 295 300

Ile Gly Pro Pro Ala Pro Asp Asn Phe Asp Tyr Leu His Glu Val Gly
305 310 315 320

Cys Lys Leu Ala His Leu Ala Glu Ala Ile His Val Glu Phe Glu Tyr
325 330 335

Arg Gly Phe Val Ala Asn Thr Leu Ala Asp Leu Asp Ala Ser Met Leu
340 345 350

Glu Leu Arg Pro Ser Glu Ile Glu Ser Val Ala Val Asn Ser Val Phe
355 360 365

Glu Leu His Lys Leu Leu Gly Arg Pro Gly Ala Ile Asp Lys Val Leu
370 375 380

Gly Val Val Asn Gln Ile Lys Pro Glu Ile Phe Thr Val Val Glu Gln
385 390 395 400

Glu Ser Asn His Asn Ser Pro Ile Phe Leu Asp Arg Phe Thr Glu Ser
405 410 415

Leu His Tyr Tyr Ser Thr Leu Phe Asp Ser Leu Glu Gly Val Pro Ser
420 425 430

Gly Gln Asp Lys Val Met Ser Glu Val Tyr Leu Gly Lys Gln Ile Cys
435 440 445

Asn Val Val Ala Cys Asp Gly Pro Asp Arg Val Glu Arg His Glu Thr
450 455 460

Leu Ser Gln Trp Arg Asn Arg Phe Gly Ser Ala Gly Phe Ala Ala Ala
465 470 475 480

His Ile Gly Ser Asn Ala Phe Lys Gln Ala Ser Met Leu Leu Ala Leu
485 490 495

Phe Asn Gly Gly Glu Gly Tyr Arg Val Glu Glu Ser Asp Gly Cys Leu
500 505 510

Met Leu Gly Trp His Thr Arg Pro Leu Ile Ala Thr Ser Ala Trp Lys
515 520 525

Leu Ser Thr Asn
530

<210> 39
<211> 1445
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (236)..(1306)

MBI-20 Sequence Listing.ST25

<223> G1944

<400> 39

```
tcgaccttcc taatttccaa cctctgttct tagcaatata ttttttctcc aaaaataatt     60

ctcagtttga ttttcttctt ctagctctta agtatatttc tttgttgtta tttatctttt    120

aatcctttaa tctcatcttt gtttatcttt aatcaaaacc caaaatttac atgggttctt    180

gaaaatctag aagaaataaa ggaaacataa caaaaataga aagaaaaaga agcta atg     238
                                                              Met
                                                               1

gtc tta aat atg gag tct acc gga gaa gct gtt aga tca acc acc ggt      286
Val Leu Asn Met Glu Ser Thr Gly Glu Ala Val Arg Ser Thr Thr Gly
            5                  10                  15

aac gac ggt ggt att acg gtg gtt aga tcc gac gcg ccg tca gat ttc      334
Asn Asp Gly Gly Ile Thr Val Val Arg Ser Asp Ala Pro Ser Asp Phe
        20                  25                  30

cac gta gct caa aga tca gaa agc tca aac caa tct ccc acc tct gtc      382
His Val Ala Gln Arg Ser Glu Ser Ser Asn Gln Ser Pro Thr Ser Val
    35                  40                  45

act cct cct cca cca cag cca tcg tct cat cac aca gct cct ccg ccg      430
Thr Pro Pro Pro Pro Gln Pro Ser Ser His His Thr Ala Pro Pro Pro
50                  55                  60                  65

ctg caa att tcg acg gtg acg act acg act acg acg gcc gcg atg gaa      478
Leu Gln Ile Ser Thr Val Thr Thr Thr Thr Thr Thr Ala Ala Met Glu
                70                  75                  80

ggt atc tcc ggt gga ctg atg aag aag aag cgt gga cgg cca agg aag      526
Gly Ile Ser Gly Gly Leu Met Lys Lys Lys Arg Gly Arg Pro Arg Lys
            85                  90                  95

tat gga ccg gac ggg act gtt gta gcg tta tct cct aaa ccg att tca      574
Tyr Gly Pro Asp Gly Thr Val Val Ala Leu Ser Pro Lys Pro Ile Ser
        100                 105                 110

tca gcg ccg gcg ccg tcg cat ctt ccg ccg ccg agt tca cac gtc atc      622
Ser Ala Pro Ala Pro Ser His Leu Pro Pro Pro Ser Ser His Val Ile
    115                 120                 125

gat ttc tcc gct tct gag aaa cgt agc aaa gtg aaa cca acg aac tcg      670
Asp Phe Ser Ala Ser Glu Lys Arg Ser Lys Val Lys Pro Thr Asn Ser
130                 135                 140                 145

ttt aac aga aca aag tat cat cac caa gtt gag aat ttg ggt gaa tgg      718
Phe Asn Arg Thr Lys Tyr His His Gln Val Glu Asn Leu Gly Glu Trp
                150                 155                 160

gct cct tgc tcc gtc ggt ggt aat ttc aca cct cat ata atc aca gtc      766
Ala Pro Cys Ser Val Gly Gly Asn Phe Thr Pro His Ile Ile Thr Val
            165                 170                 175

aac acc ggc gag gat gta aca atg aag ata atc tcg ttt tcg caa caa      814
Asn Thr Gly Glu Asp Val Thr Met Lys Ile Ile Ser Phe Ser Gln Gln
        180                 185                 190

gga cct cgc tct att tgt gtt ctg tca gca aac ggt gtt att tca agc      862
Gly Pro Arg Ser Ile Cys Val Leu Ser Ala Asn Gly Val Ile Ser Ser
    195                 200                 205

gtt aca ctt cgt cag cca gat tcc tct ggc ggc aca ttg aca tac gaa      910
Val Thr Leu Arg Gln Pro Asp Ser Ser Gly Gly Thr Leu Thr Tyr Glu
210                 215                 220                 225

ggt cgg ttt gag ata tta tca tta tcc ggg tca ttc atg cct aat gat      958
Gly Arg Phe Glu Ile Leu Ser Leu Ser Gly Ser Phe Met Pro Asn Asp
                230                 235                 240

tca ggc gga aca cga agt aga acg gga gga atg agt gta tcg tta gca     1006
```

MBI-20 Sequence Listing.ST25

```
Ser Gly Gly Thr Arg Ser Arg Thr Gly Gly Met Ser Val Ser Leu Ala
            245                 250                 255

agt ccc gat gga cgt gta gta ggc ggt ggc ctc gcc ggt tta cta gta     1054
Ser Pro Asp Gly Arg Val Val Gly Gly Gly Leu Ala Gly Leu Leu Val
        260                 265                 270

gcc gcg agt ccg gtt cag gtg gtt gta gga agt ttt tta gcg ggc act     1102
Ala Ala Ser Pro Val Gln Val Val Val Gly Ser Phe Leu Ala Gly Thr
    275                 280                 285

gac cat caa gat cag aaa ccg aaa aag aac aaa cat gat ttc atg ttg     1150
Asp His Gln Asp Gln Lys Pro Lys Lys Asn Lys His Asp Phe Met Leu
290                 295                 300                 305

tcg agt cct acc gct gca att cct atc tct agt gca gct gat cac cgg     1198
Ser Ser Pro Thr Ala Ala Ile Pro Ile Ser Ser Ala Ala Asp His Arg
                310                 315                 320

aca atc cat tcg gtc tcg tct ctt ccg gtc aat aat aat aca tgg cag     1246
Thr Ile His Ser Val Ser Ser Leu Pro Val Asn Asn Asn Thr Trp Gln
            325                 330                 335

act tct tta gct tcc gat cca aga aac aag cat acc gat att aat gtc     1294
Thr Ser Leu Ala Ser Asp Pro Arg Asn Lys His Thr Asp Ile Asn Val
        340                 345                 350

aat gta act tga aatccaatct ttctctgtat tttctgttaa caagtttgat         1346
Asn Val Thr
    355

ttggttgttt atctacatta ggattttact aaaatggtag tattatttat agggttttag   1406

ggtctttatt ttggttccac tgttgtcact tgtaggata                          1445


<210>  40
<211>  356
<212>  PRT
<213>  Arabidopsis thaliana

<400>  40

Met Val Leu Asn Met Glu Ser Thr Gly Glu Ala Val Arg Ser Thr Thr
1               5                   10                  15

Gly Asn Asp Gly Gly Ile Thr Val Val Arg Ser Asp Ala Pro Ser Asp
            20                  25                  30

Phe His Val Ala Gln Arg Ser Glu Ser Ser Asn Gln Ser Pro Thr Ser
        35                  40                  45

Val Thr Pro Pro Pro Pro Gln Pro Ser Ser His His Thr Ala Pro Pro
    50                  55                  60

Pro Leu Gln Ile Ser Thr Val Thr Thr Thr Thr Thr Thr Ala Ala Met
65                  70                  75                  80

Glu Gly Ile Ser Gly Gly Leu Met Lys Lys Lys Arg Gly Arg Pro Arg
                85                  90                  95

Lys Tyr Gly Pro Asp Gly Thr Val Val Ala Leu Ser Pro Lys Pro Ile
            100                 105                 110

Ser Ser Ala Pro Ala Pro Ser His Leu Pro Pro Pro Ser Ser His Val
        115                 120                 125
```

MBI-20 Sequence Listing.ST25

Ile Asp Phe Ser Ala Ser Glu Lys Arg Ser Lys Val Lys Pro Thr Asn
130 135 140

Ser Phe Asn Arg Thr Lys Tyr His His Gln Val Glu Asn Leu Gly Glu
145 150 155 160

Trp Ala Pro Cys Ser Val Gly Gly Asn Phe Thr Pro His Ile Ile Thr
165 170 175

Val Asn Thr Gly Glu Asp Val Thr Met Lys Ile Ile Ser Phe Ser Gln
180 185 190

Gln Gly Pro Arg Ser Ile Cys Val Leu Ser Ala Asn Gly Val Ile Ser
195 200 205

Ser Val Thr Leu Arg Gln Pro Asp Ser Ser Gly Gly Thr Leu Thr Tyr
210 215 220

Glu Gly Arg Phe Glu Ile Leu Ser Leu Ser Gly Ser Phe Met Pro Asn
225 230 235 240

Asp Ser Gly Gly Thr Arg Ser Arg Thr Gly Gly Met Ser Val Ser Leu
245 250 255

Ala Ser Pro Asp Gly Arg Val Val Gly Gly Gly Leu Ala Gly Leu Leu
260 265 270

Val Ala Ala Ser Pro Val Gln Val Val Val Gly Ser Phe Leu Ala Gly
275 280 285

Thr Asp His Gln Asp Gln Lys Pro Lys Lys Asn Lys His Asp Phe Met
290 295 300

Leu Ser Ser Pro Thr Ala Ala Ile Pro Ile Ser Ser Ala Ala Asp His
305 310 315 320

Arg Thr Ile His Ser Val Ser Ser Leu Pro Val Asn Asn Asn Thr Trp
325 330 335

Gln Thr Ser Leu Ala Ser Asp Pro Arg Asn Lys His Thr Asp Ile Asn
340 345 350

Val Asn Val Thr
355

<210> 41
<211> 1558
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (191)..(1396)
<223> G326

<400> 41
caattaatga catcttcttc ttctcctttc actgcaaaac cgaaagcttg agactttgag 60

attatgtcta tgtcatcttc ttcttcttcc atcgatcact tcatcacctt tcgtcatctt 120

MBI-20 Sequence Listing.ST25

```
gatcttattc tccactgtat aaaatcagcg agattttaag ggattgtgaa ggtaccatct    180

taaacacaaa atg ggt act tct act aca gag agt gtg gtg gcg tgt gaa       229
           Met Gly Thr Ser Thr Thr Glu Ser Val Val Ala Cys Glu
           1               5                   10

ttt tgc ggc gag aga acg gcg gtt ctg ttt tgt aga gcc gat acg gcg      277
Phe Cys Gly Glu Arg Thr Ala Val Leu Phe Cys Arg Ala Asp Thr Ala
    15                  20                  25

aag ctt tgt ttg cct tgt gac cag cac gtg cac tcg gcg aac ctt ctc      325
Lys Leu Cys Leu Pro Cys Asp Gln His Val His Ser Ala Asn Leu Leu
30                  35                  40                  45

tcg agg aag cat gtt cgt tct cag atc tgt gat aac tgt agc aaa gag      373
Ser Arg Lys His Val Arg Ser Gln Ile Cys Asp Asn Cys Ser Lys Glu
                50                  55                  60

ccg gtg tcc gta cgt tgc ttc aca gat aat ctc gta ttg tgt cag gag      421
Pro Val Ser Val Arg Cys Phe Thr Asp Asn Leu Val Leu Cys Gln Glu
            65                  70                  75

tgt gat tgg gat gtt cac gga agc tgt tcc tcc tcc gcg acg cat gaa      469
Cys Asp Trp Asp Val His Gly Ser Cys Ser Ser Ser Ala Thr His Glu
        80                  85                  90

cgc tcc gcc gtg gaa ggg ttt tca ggt tgt cct tcg gtt ttg gag ctt      517
Arg Ser Ala Val Glu Gly Phe Ser Gly Cys Pro Ser Val Leu Glu Leu
    95                  100                 105

gct gct gtg tgg gga atc gat tta aag ggt aag aag aaa gaa gat gac      565
Ala Ala Val Trp Gly Ile Asp Leu Lys Gly Lys Lys Lys Glu Asp Asp
110                 115                 120                 125

gaa gac gaa ttg act aag aat ttt ggg atg ggg ttg gat tcg tgg ggt      613
Glu Asp Glu Leu Thr Lys Asn Phe Gly Met Gly Leu Asp Ser Trp Gly
                130                 135                 140

tct gga tct aac atc gtt caa gaa ctg att gtt cct tat gat gtg tct      661
Ser Gly Ser Asn Ile Val Gln Glu Leu Ile Val Pro Tyr Asp Val Ser
            145                 150                 155

tgc aaa aag caa agc ttt agc ttt ggg agg tct aag cag gta gtg ttt      709
Cys Lys Lys Gln Ser Phe Ser Phe Gly Arg Ser Lys Gln Val Val Phe
        160                 165                 170

gaa cag ctt gag tta ctg aag aga ggc ttc gtt gaa ggc gaa gga gag      757
Glu Gln Leu Glu Leu Leu Lys Arg Gly Phe Val Glu Gly Glu Gly Glu
    175                 180                 185

att atg gtt ccg gag gga atc aat ggc gga gga agc att tct cag cca      805
Ile Met Val Pro Glu Gly Ile Asn Gly Gly Gly Ser Ile Ser Gln Pro
190                 195                 200                 205

tct ccg acg acg tcg ttt act tct ttg ctt atg tct caa agt ctt tgt      853
Ser Pro Thr Thr Ser Phe Thr Ser Leu Leu Met Ser Gln Ser Leu Cys
                210                 215                 220

ggt aat ggt atg caa tgg aat gct act aat cat agc act ggc cag aac      901
Gly Asn Gly Met Gln Trp Asn Ala Thr Asn His Ser Thr Gly Gln Asn
            225                 230                 235

act cag ata tgg gat ttt aac ttg gga cag tcg agg aac cct gat gaa      949
Thr Gln Ile Trp Asp Phe Asn Leu Gly Gln Ser Arg Asn Pro Asp Glu
        240                 245                 250

cct agt cca gtc gaa act aaa ggc tct act ttc aca ttc aac aac gtt      997
Pro Ser Pro Val Glu Thr Lys Gly Ser Thr Phe Thr Phe Asn Asn Val
    255                 260                 265

act cat ctc aag aac gat acc cga acc acc aat atg aat gct ttc aaa     1045
Thr His Leu Lys Asn Asp Thr Arg Thr Thr Asn Met Asn Ala Phe Lys
270                 275                 280                 285
```

MBI-20 Sequence Listing.ST25

```
gag agt tac cag gag gat tcc gtc cac tca act tct acc aag gga cag     1093
Glu Ser Tyr Gln Glu Asp Ser Val His Ser Thr Ser Thr Lys Gly Gln
                290                 295                 300

gaa aca tct aag agc aac aat att cct gct gcc att cac tcg cat aaa     1141
Glu Thr Ser Lys Ser Asn Asn Ile Pro Ala Ala Ile His Ser His Lys
            305                 310                 315

agt tct aac gac tcc tgt ggc ttg cat tgc acg gaa cat att gct att     1189
Ser Ser Asn Asp Ser Cys Gly Leu His Cys Thr Glu His Ile Ala Ile
        320                 325                 330

act agt aat aga gcc aca aga ttg gtg gcg gta acg aat gct gat cta     1237
Thr Ser Asn Arg Ala Thr Arg Leu Val Ala Val Thr Asn Ala Asp Leu
    335                 340                 345

gag cag atg gca cag aac aga gat aat gct atg cag cgg tac aag gaa     1285
Glu Gln Met Ala Gln Asn Arg Asp Asn Ala Met Gln Arg Tyr Lys Glu
350                 355                 360                 365

aag aag aaa acg cgg aga tat gat aag acc ata aga tat gaa acg agg     1333
Lys Lys Lys Thr Arg Arg Tyr Asp Lys Thr Ile Arg Tyr Glu Thr Arg
                370                 375                 380

aag gcg aga gcc gag acc agg ttg cgt gtt aag ggc aga ttt gtg aaa     1381
Lys Ala Arg Ala Glu Thr Arg Leu Arg Val Lys Gly Arg Phe Val Lys
            385                 390                 395

gct aca gat cct tag atgtctctcc acgttaggtt ttacatttga gatcctaagt     1436
Ala Thr Asp Pro
        400

taggaacttt ttttgttttt tctactttca actaccttgt aaatgtaaat gatcgatctt   1496

cagctgcata atgtgtggcc agatttttgt aatttttacg tttaaccttc taaaaaaaaa   1556

aa                                                                  1558
```

<210> 42
<211> 401
<212> PRT
<213> Arabidopsis thaliana

<400> 42

```
Met Gly Thr Ser Thr Thr Glu Ser Val Val Ala Cys Glu Phe Cys Gly
1               5                   10                  15

Glu Arg Thr Ala Val Leu Phe Cys Arg Ala Asp Thr Ala Lys Leu Cys
            20                  25                  30

Leu Pro Cys Asp Gln His Val His Ser Ala Asn Leu Leu Ser Arg Lys
        35                  40                  45

His Val Arg Ser Gln Ile Cys Asp Asn Cys Ser Lys Glu Pro Val Ser
    50                  55                  60

Val Arg Cys Phe Thr Asp Asn Leu Val Leu Cys Gln Glu Cys Asp Trp
65                  70                  75                  80

Asp Val His Gly Ser Cys Ser Ser Ser Ala Thr His Glu Arg Ser Ala
                85                  90                  95

Val Glu Gly Phe Ser Gly Cys Pro Ser Val Leu Glu Leu Ala Ala Val
            100                 105                 110

Trp Gly Ile Asp Leu Lys Gly Lys Lys Lys Glu Asp Asp Glu Asp Glu
```

115 120 125

Leu Thr Lys Asn Phe Gly Met Gly Leu Asp Ser Trp Gly Ser Gly Ser
130 135 140

Asn Ile Val Gln Glu Leu Ile Val Pro Tyr Asp Val Ser Cys Lys Lys
145 150 155 160

Gln Ser Phe Ser Phe Gly Arg Ser Lys Gln Val Val Phe Glu Gln Leu
165 170 175

Glu Leu Leu Lys Arg Gly Phe Val Glu Gly Glu Gly Glu Ile Met Val
180 185 190

Pro Glu Gly Ile Asn Gly Gly Gly Ser Ile Ser Gln Pro Ser Pro Thr
195 200 205

Thr Ser Phe Thr Ser Leu Leu Met Ser Gln Ser Leu Cys Gly Asn Gly
210 215 220

Met Gln Trp Asn Ala Thr Asn His Ser Thr Gly Gln Asn Thr Gln Ile
225 230 235 240

Trp Asp Phe Asn Leu Gly Gln Ser Arg Asn Pro Asp Glu Pro Ser Pro
245 250 255

Val Glu Thr Lys Gly Ser Thr Phe Thr Phe Asn Asn Val Thr His Leu
260 265 270

Lys Asn Asp Thr Arg Thr Thr Asn Met Asn Ala Phe Lys Glu Ser Tyr
275 280 285

Gln Glu Asp Ser Val His Ser Thr Ser Thr Lys Gly Gln Glu Thr Ser
290 295 300

Lys Ser Asn Asn Ile Pro Ala Ala Ile His Ser His Lys Ser Ser Asn
305 310 315 320

Asp Ser Cys Gly Leu His Cys Thr Glu His Ile Ala Ile Thr Ser Asn
325 330 335

Arg Ala Thr Arg Leu Val Ala Val Thr Asn Ala Asp Leu Glu Gln Met
340 345 350

Ala Gln Asn Arg Asp Asn Ala Met Gln Arg Tyr Lys Glu Lys Lys Lys
355 360 365

Thr Arg Arg Tyr Asp Lys Thr Ile Arg Tyr Glu Thr Arg Lys Ala Arg
370 375 380

Ala Glu Thr Arg Leu Arg Val Lys Gly Arg Phe Val Lys Ala Thr Asp
385 390 395 400

Pro

MBI-20 Sequence Listing.ST25

<210> 43
<211> 844
<212> DNA
<213> Arabidopsis thaliana

<220>
<221> CDS
<222> (89)..(658)
<223> G1387

<400> 43

```
tctctctccc actctcactt tctctcctat tcttagttcg tgtcagaaac acacagagaa      60

attaagaacc ctaatttaaa acagaaga atg gta cat tcg aag aag ttc cga         112
                                  Met Val His Ser Lys Lys Phe Arg
                                  1               5

ggt gtc cgc cag cgt cag tgg ggt tct tgg gtt tct gag att cgt cat         160
Gly Val Arg Gln Arg Gln Trp Gly Ser Trp Val Ser Glu Ile Arg His
    10                  15                  20

cct ctc ttg aag aga aga gtg tgg cta gga aca ttc gac acg gcg gaa         208
Pro Leu Leu Lys Arg Arg Val Trp Leu Gly Thr Phe Asp Thr Ala Glu
25                  30                  35                  40

aca gcg gct aga gcc tac gac caa gcc gcg gtt cta atg aac ggc cag         256
Thr Ala Ala Arg Ala Tyr Asp Gln Ala Ala Val Leu Met Asn Gly Gln
                45                  50                  55

agc gcg aag act aac ttc ccc gtc atc aaa tcg aac ggt tca aat tcc         304
Ser Ala Lys Thr Asn Phe Pro Val Ile Lys Ser Asn Gly Ser Asn Ser
            60                  65                  70

ttg gag att aac tct gcg tta agg tct ccc aaa tca tta tcg gaa cta         352
Leu Glu Ile Asn Ser Ala Leu Arg Ser Pro Lys Ser Leu Ser Glu Leu
        75                  80                  85

ttg aac gct aag cta agg aag aac tgt aaa gac cag aca ccg tat ctg         400
Leu Asn Ala Lys Leu Arg Lys Asn Cys Lys Asp Gln Thr Pro Tyr Leu
    90                  95                  100

acg tgt ctc cgc ctc gac aac gac agc tca cac atc ggc gtc tgg cag         448
Thr Cys Leu Arg Leu Asp Asn Asp Ser Ser His Ile Gly Val Trp Gln
105                 110                 115                 120

aaa cgc gcc ggg tca aaa acg agt cca aac tgg gtc aag ctt gtt gaa         496
Lys Arg Ala Gly Ser Lys Thr Ser Pro Asn Trp Val Lys Leu Val Glu
                125                 130                 135

cta ggt gac aaa gtt aac gca cgt ccc ggt ggt gat att gag act aat         544
Leu Gly Asp Lys Val Asn Ala Arg Pro Gly Gly Asp Ile Glu Thr Asn
            140                 145                 150

aag atg aag gta cga aac gaa gac gtt cag gaa gat gat caa atg gcg         592
Lys Met Lys Val Arg Asn Glu Asp Val Gln Glu Asp Asp Gln Met Ala
        155                 160                 165

atg cag atg atc gag gag ttg ctt aac tgg acc tgt cct gga tct gga         640
Met Gln Met Ile Glu Glu Leu Leu Asn Trp Thr Cys Pro Gly Ser Gly
    170                 175                 180

tcc att gca cag gtc taa aggagaatca ttgaattata tgatcaagat               688
Ser Ile Ala Gln Val
185

aataatatag ttgagggtta ataataatcg agggtaagta atttacgtgt agctaataat      748

taatataatt ttcgaacata tatatgaata tatgatagct ctagaaatga gtacgtatat      808

atacgtaaac atttttcctc aaatatagta tatgtg                                844
```

<210> 44
<211> 189

MBI-20 Sequence Listing.ST25

<212> PRT
<213> Arabidopsis thaliana

<400> 44

Met Val His Ser Lys Lys Phe Arg Gly Val Arg Gln Arg Gln Trp Gly
1 5 10 15

Ser Trp Val Ser Glu Ile Arg His Pro Leu Leu Lys Arg Arg Val Trp
20 25 30

Leu Gly Thr Phe Asp Thr Ala Glu Thr Ala Ala Arg Ala Tyr Asp Gln
35 40 45

Ala Ala Val Leu Met Asn Gly Gln Ser Ala Lys Thr Asn Phe Pro Val
50 55 60

Ile Lys Ser Asn Gly Ser Asn Ser Leu Glu Ile Asn Ser Ala Leu Arg
65 70 75 80

Ser Pro Lys Ser Leu Ser Glu Leu Leu Asn Ala Lys Leu Arg Lys Asn
85 90 95

Cys Lys Asp Gln Thr Pro Tyr Leu Thr Cys Leu Arg Leu Asp Asn Asp
100 105 110

Ser Ser His Ile Gly Val Trp Gln Lys Arg Ala Gly Ser Lys Thr Ser
115 120 125

Pro Asn Trp Val Lys Leu Val Glu Leu Gly Asp Lys Val Asn Ala Arg
130 135 140

Pro Gly Gly Asp Ile Glu Thr Asn Lys Met Lys Val Arg Asn Glu Asp
145 150 155 160

Val Gln Glu Asp Asp Gln Met Ala Met Gln Met Ile Glu Glu Leu Leu
165 170 175

Asn Trp Thr Cys Pro Gly Ser Gly Ser Ile Ala Gln Val
180 185

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US00/31344

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12N 5/04, 5/10, 15/00, 15/09, 15/63, 15/70, 15/74, 15/82, 15/87; C07H 21/02, 21/04; A01H 1/00, 9/00, 11/00

US CL : 435/320.1, 419, 468; 536/23.1; 800/ 278, 295

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/320.1, 419, 468; 536/23.1; 800/ 278, 295

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	Database GenEmbl, Accession Number U28422, WANG et al., Arabidopsis thaliana Dna-binding protein CCAL (CCAL) mRNA, complete cds 14 January 1997.	4-6 --- 1-3, 7-13, 25-27
X --- Y	Database Geneseq., Accession Number V65382, THE REGENTS OF THE UNIVERSITY OF CALIFORNIA. 15 February 1999.	4-6 --- 1-3, 7-13, 25-27
X --- Y	Database PIR. Accession Number T02684, ROUNSLEY et al., DNA-binding protein CCA1 - Arabidopsis thaliana. 24 March 1999.	11 --- 1-10, 12-13, 25-27
X --- Y	Database Geneseq, Accession Number W79280,THE REGENTS OF THE UNIVERSITY OF CALIFORNIA. 15 February 1999.	11 --- 1-10, 12-13, 25-27
X	WO 98/48007 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA) 28 October 1998, pages 26-33, pages 43-46 SEQ ID NO: 3.	1-13, 25-27
P,Y	RIECHMANN et al. A genomic perspective on plant transcription factors. Current Opinion in Plant Biology. October 2000, Vol. 3, No. 5, pages 423-434, especially pages 427-428.	1-13, 25-27

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
08 February 2001 (08.02.2001)	07 MAR 2001
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Cynthia Collins Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/31344

C (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	RIECHMANN et al. The AP2/EREBP family of plant transcription factors. Biological Chemistry. June 1998. Vol. 379. No. 6. pages 633-646.	1-13. 25-27
A	RIECHMANN et al. MADS domain proteins in plant development. Biological Chemistry. October 1997. Vol. 378. No. 10. pages 1079-1101.	1-13. 25-27
A	HEARD et al. Evolutionary diversity of symbiotically induced nodule MADS box genes: characterization of nmhC5. a member of a novel subfamily. Molecular plant-microbe interactions: MPMI. July 1997. Vol. 10. No. 5. pages 665-676.	1 13. 25-27
A	HEARD et al. Symbiotic induction of a MADS-box gene during development of alfalfa root nodules. Proc. Natl. Acad. Sci. USA. 06 June 1995. Vol. 92. No. 12. pages 5273-5277.	1-13. 25-27

INTERNATIONAL SEARCH REPORT

:rnational application No.
PCT/US00/31344

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claim Nos.: 14
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.: 1-13, 25-27 SEQ ID NOS:1 and 2

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US00/31344

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Groups I-XXII, claim(s) 1-13 and 25-27, drawn to transgenic plants with modified biochemical characteristics, polynucleotides and vectors for producing said transgenic plants, and methods of making said transgenic plants. Applicant must elect one pair of sequences (one nucleotide sequence and its corresponding amino acid translation) per Group to be examined, *i.e.* SEQ ID NOS: 1 and 2 in Group I, SEQ ID NOS: 3 and 4 in Group II, SEQ ID NOS: 5 and 6 in Group III, etc.

Group XXIII, claim(s) 15-17, drawn to a method of identifying a factor that is modulated by or interacts with a polypeptide.

Group XXIV, claim(s) 18, drawn to a method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide of interest.

Group XXV, claim(s) 19 and 20, drawn to an integrated system, computer, or computer readable medium.

Group XXVI, claim(s) 21-23, drawn to a method of identifying a polynucleotide sequence.
The inventions listed as Groups I-XXVI do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

The inventions listed as Groups I-XXVI do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: Groups I-XXII are drawn to transgenic plants and methods of producing said plants with nucleic acid sequences. The methods of Groups I-XXII differ from each other in that they are directed to plant transformation methods and transgenic plants with structurally and functionally distinct nucleic acid sequences which encode structurally and functionally different amino acid sequences. In addition, Groups XXIII, XXIV, and XXVI are different methods from any of Groups I-XXII in that they have different method steps and different end products, and Group XXV requires a computer system. Thus, there is no single special technical feature which links the inventions of Groups I-XXVI under PCT Rule 13.2.

Continuation of B. FIELDS SEARCHED Item 3: STN (agricola, biosis, biotechno, biotechds, biotechabs, caba, caplus, embase, medline,uspatfull, wpids, pctfull, europatfull, japio) SEARCH TERMS: inventor names, plant transcription factor, fatty acid, chlorophyll, carotenoid; STIC sequence search for SEQ ID NOS: 1 and 2